Studies towards the prebiotic synthesis of nucleotides and amino acids

Kathryn Ashe

UCL

2018

Thesis submitted to University College London
for the degree of Doctor of Philosophy

UCL Department of Chemistry
20 Gordon Street
London
WC1H 0AJ

I, Kathryn Ashe, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Understanding how life on Earth might have started is the major goal of prebiotic chemistry. The universal nature of extant life, with integrated informational, catalytic and compartment-forming systems, suggests a common chemical origin. Comprehensive studies of the feasible routes towards the biomolecules necessary for life will refine our knowledge of the prebiotic chemistry that lead to life.

The integration of genetic information and proteins in the central dogma of molecular biology indicates the importance of evaluating the prebiotic synthesis of amino acids and peptides. Following the principles of prebiotic systems chemistry, ideally this should take place under similar conditions established for other prebiotically plausible networks that lead to other key class of metabolite using common reagents. Here we focus on resolving the pH incompatibility of nucleotide (pH <7) and amino acid synthesis (by Strecker reaction of ammonium cyanide and aldehydes, pH >9). Introducing DAP – a reagent previously implicated in various areas of prebiotic chemistry – into the Strecker reaction enables the synthesis of *N*-phosphoroaminonitriles at pH 7, and is highly selective for proteinogenic over non-proteinogenic amino acid precursors. Derivatisation of the products of this phosphoro-Strecker reaction provides a handle for further reaction and possible oligomerisation. Phosphate is used as a universal energy currency in extant life, notably in amino acid activation in protein synthesis, so the formation of a phosphorylated amino acid derivative in this work is particularly interesting.

The ubiquity of RNA in modern biochemistry and RNA's dual functionality (genotypic and phenotypic) makes RNA a strong contender for the first biopolymer. We also present an account of our results exploring: the replacement of phosphate with borate in the stereochemical rearrangement of 2′,2-anhydropyrimidines and 2′,8-anhydropurines; the formation of a glycosidic bond between acetylated ribose derivatives (from the UV irradiation of unwanted α-pyrimidine by-products) and purine nucleobases; and a stereospecific reduction of 5′,8-cyclopurines.

Impact statement

A systems chemistry approach to investigating the origin of life requires the experimental evaluation of the chemistry that could have occurred on the early Earth, to discover both what could have occurred and what was unlikely to occur, and use commonality of substrates, intermediates, by-products and conditions to not only link disparate areas of prebiotic synthesis but also to drive new discoveries and hypotheses. As such, the research presented here forms a small part of the puzzle that allows us to gain a greater understanding of the processes that governed the advent of life.

The insights revealed by this work will be primarily of interest to the origin of life research community, disseminated through journal articles and conference talks. This project has revealed a new route towards amino acid precursors by consideration of prebiotic nucleotide chemistry, indicating a need for future research to explore the full chemical potential of these precursors. Furthermore, this synthesis is accomplished using a reagent (DAP) that has already been shown to effect a number of transformations in other areas of prebiotic chemistry, and therefore emphasises the benefits of the continued investigation of the use of systems chemical analysis in prebiotic chemistry. The work presented here also demonstrates transformations that, although appealing in a prebiotic scenario, were unlikely to have occurred on the early Earth, and therefore provides parameters to constrain future research. These findings, in conjunction with other origins of life research, represent one of many incremental steps towards developing an overview of the chemistry that led to the start of life.

The question of our ultimate origin is one that has captured the imagination of the public for generations. Public engagement activities, based on the combined insights produced by origin of life research, can enthuse members of the public and inform their understanding of both our current understanding of life's origin and, more generally, the principles of scientific research. Origin of life research can be used as a particularly good case study in the latter respect: it is an interdisciplinary area, offering an example of how scientific research today and in the future increasingly relies on bringing together knowledge and insights from disparate subjects to generate an overview of a subject. The author has presented lay-friendly talks at university open days, with the objective of informing and engaging prospective students and their parents in scientific research generally, and origin of life research in particular, with this objective in mind.

As well as raising the awareness of the general public, stimulating discussion and debate, outreach activity targeted at schools will inspire young people to study sciences, ensuring the future continuity of research into this fascinating area.

Contents

Acknowledgements

Firstly, I would like to thank my supervisor, Matt Powner, for giving me the opportunity to carry out research in such an interesting area. I would also like to thank all members of the Powner group both past and present for many hours of discussion and help in and out of the lab, and particularly Adam, Arif, Christian, Natalie and Shaun for their helpful suggestions. My thanks also to Becca, Cal, Daniel, Lello and Sam for proof-reading this thesis and offering corrections.

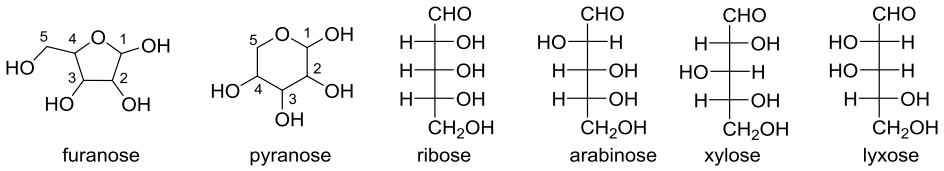
And, finally, I would like to thank my friends and family, without whose support this would have been impossible – and especially my parents, for always pushing me to discover more.

Abbreviations

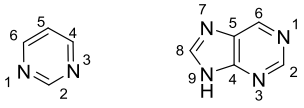
A	adenine
Ac	acetyl
Acetyl-CoA	acetyl coenzyme A
AICA	4-amino-imidazole-5-carboxamide
AICN	4-amino-imidazole-5-carbonitrile
aq.	aqueous
AmTP	amidotriphosphate
ATP	adenosine triphosphate
Bz	benzoyl
C	cytosine
°C	degrees Celsius
calcd	calculated
COS	carbonyl sulfide
COSY	correlated spectroscopy
DAMN	diaminomaleonitrile
DAP	diamidophosphate
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
Et	ethyl
et al	*et alia* (Latin: and others)
G	guanine
Ga	giga annum (billion years)
Gya	billion years ago
h	hour(s)
HMBC	heteronuclear multi-bond correlation
HSQC	heteronuclear single quantum correlation
Hz	Hertz
IR	infrared
J	NMR coupling constant (measured in Hz)
LUCA	last universal common ancestor
M	molar
Me	methyl
MHz	mega Hertz
m.p.	melting point
MS	mass spectroscopy
m/z	mass/charge ratio
NCA	*N*-carboxyanhydride
NMR	nuclear magnetic resonance
Ph	phenyl
Pi	orthophosphate, sodium ion (unless otherwise stated)
PPi	pyrophosphate
RNA	ribonucleic acid
rt	room temperature
U	uracil
UV	ultraviolet
XNA	xeno nucleic acid

Numbering and nomenclature

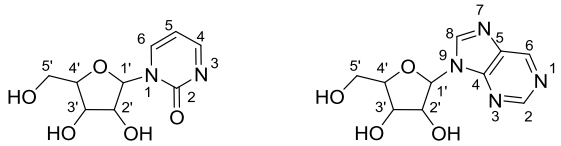
pentose sugars



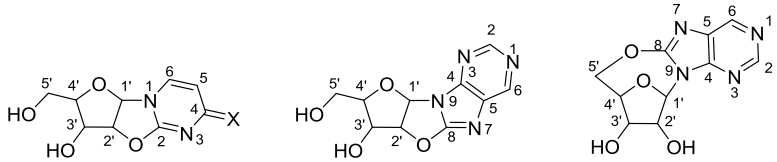
nucleobases: pyrimidines (left); purines (right)



nucleosides



anhydronucleosides



1 Introduction

The distinguishing feature of Earth is the presence of life. Earth is uniquely the only place throughout the universe known to support life, and indeed does so to an astounding extent: almost every environment on Earth is colonised by life, such that life has changed the characteristics of Earth's atmosphere, hydrosphere and climate. Part of the wider scientific inquiry into understanding the operation and function of the universe is to understand life (both what it is and how it works), and therefore also how life began.

1.1 What is life?

Life is characterised by the presence of catalytic and genetic polymers in a steady state far from equilibrium;[1] the simplest hypothetical cell can be reduced to informational, catalytic and compartment-forming systems linked by metabolism, the network of chemical reactions providing the materials and energy required for cellular processes (***Figure 1.1***).[2] Life exists within an environment, and so the properties and function of a single organism cannot be considered in isolation from its environmental context. Indeed, some researchers even make the distinction between 'being alive' (a transient state of being allowed by metabolic activity) and 'life' (the interaction of all living organisms with the environment which has existed since its origin).[3]

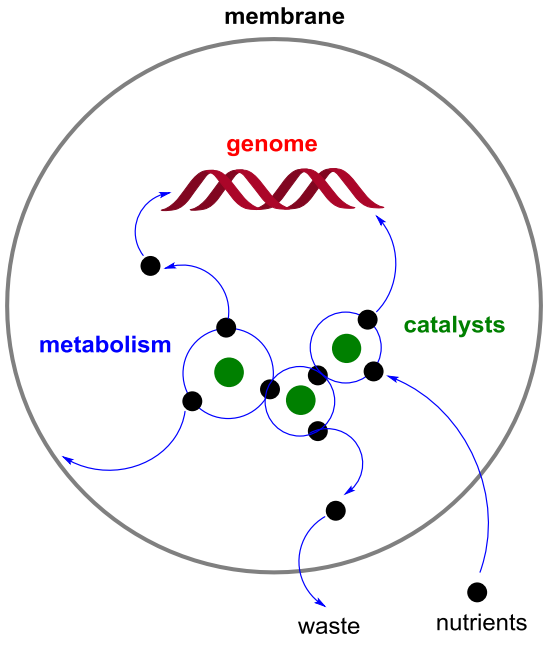


Figure 1.1*: Conceptual reduction to a minimal cell containing genomic, catalytic and compartment-forming systems linked by metabolism.*

The central dogma of molecular biology (***Figure 1.2***) describes the flow of genetic information, from storage to function, that underpins information transfer in all forms of life present. Understanding how this flow of information started is critical to understanding the origin of life. Therefore, the origin of life can be simply defined as a point in time when a subset of polymers (which had specific physical and chemical properties produced by energy flow through simple systems of molecules) was captured in a compartment and interacted to support self-sustaining replication.[1]

Figure 1.2*: The central dogma of molecular biology describes the general transfer of genetic information in a biological system. The information held in DNA (the genotype) is transcribed into RNA, and then translated into protein. The functions carried out by proteins contribute to the properties of a cell: its phenotype.*

Research into life, and its origins, is a multidisciplinary venture, but life itself is based on networks of chemical reactions, and appears to be a natural consequence of the specific organisation of chemical matter.[4] Prebiotic chemistry – the investigations of the reactions that took place on the early Earth before, but enabling, the advent of biology – is therefore key to understanding the origin of life. The number of possible reactants and environmental conditions on the early Earth, and the seemingly infinite permutations of chemical reactions that could arise from these, appear to create an insurmountable task for the prebiotic chemist.

Researchers have approached the problem of how life started using different strategies to simplify and constrain the scope of this task, which, although attractive in theory, lead to further problems in practice. A biological 'top down' perspective *via* retrodictive analysis of life today back through evolution towards increasingly simpler organisms seeks to understand the biochemistry of the earliest organism, so that the biomolecules required of prebiotic synthesis can be inferred. The geochemical 'bottom up' approach aspires to understand the conditions and materials available on the early Earth, to inform and constrain the synthesis of molecules upon which life is based. In addition, some researchers favour 'hands off' studies, in which the researcher only interferes with the experiment to introduce fresh starting materials and to analyse the products;

others the 'hands on' approach, using familiar synthetic chemistry to prove that biological molecules can be formed in prebiotic conditions.

1.2 Geochemical, geological and biological evidence

The Earth formed just over 4.5 Gya,[5] but a period of heavy asteroid bombardment until approximately 3.8 Gya would have affected the ability of any life formed to survive. Large impacts, which almost certainly took place up to 4.2 Gya and perhaps until 3.8 Gya, would have vaporised the oceans (water had been present from when Earth formed[6]) and sterilised the planet, meaning that even if life had started before then, it was likely not the ancestor of the life present on Earth today.[7] Studies of molecules observed in interstellar gases show that many organic molecules can be found in space, and therefore we can assume that these molecules were also present 4.5 Gya (***Figure 1.3***).[8] The formation of the Earth was perhaps violent enough to destroy these molecules in the surrounding environment; on the other hand, enough material may have survived to influence chemical evolution of the primitive Earth.[8]

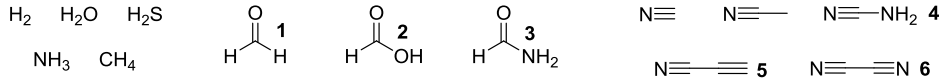


Figure 1.3*: Partial repertoire of assumed prebiotically plausible compounds, including formaldehyde* ***1****, formic acid* ***2****, formamide* ***3****, cyanamide* ***4****, cyanoacetylene* ***5,*** *and cyanogen* ***6.***

The period following this time, during which life originated, probably had a 'weakly reduced' atmosphere, consisting mostly of carbon dioxide and nitrogen, with carbon monoxide, hydrogen and reduced sulfur gases also present, and an aqueous ocean.[7] Geochemical evidence, and the fact that all organisms on Earth are able to metabolise ammonia, suggest that ammonia was also present in the primitive atmosphere.[9] The concentration of oxygen would have been low until the evolution of photosynthesis, although there are conflicting reports as to when this was: some geochemical studies show that the concentration of oxygen reached an appreciable level approximately 2 Gya;[7] others indicate it was rather earlier – approximately 3 Gya.[10]

The composition of the primordial atmosphere has a substantial influence on the UV environment.[11,12] The prebiotic environment would have experienced higher levels of UV radiation compared to today due to a greater variance in emission from the young Sun and the lack of UV-shielding (biogenic) O_2 and O_3 in the Earth's atmosphere.[12] Much of the variance in emission from the young Sun would have been as shorter

wavelengths, which would have been robustly blocked by even low concentrations of water (λ <168 nm) and carbon dioxide (λ < 204 nm) in the atmosphere.[12] Methane, O_2 and O_3 present in the atmosphere would also theoretically be capable of blocking UV radiation, but this shielding is negligible at the likely concentrations of carbon dioxide present.[11] Hydrogen cyanide and ethane are sensitive to UV photolysis, but this chemistry would largely be removed by carbon dioxide shielding.[12] Indeed, as such a low concentration of carbon dioxide and water are required to act as effective UV-shields, prebiotic photochemistry would be insensitive to the precise levels of UV-shielding compounds in the atmosphere.[11] It is highly likely that the exact UV environment would be local, changing with latitude, surface type and local atmospheric composition:[11] terrestrial UV levels would differ as a result of albedo and zenith angle, and would be sensitive to high atmospheric levels of SO_2 and H_2S in locations and times of high volcanism.

The geochemical evidence available does little to constrain the focus of researchers to one particular scenario in which life must have arisen: the early Earth would have offered a wide range of environments and conditions, and the systematic exploration of the chemistry of all of them is virtually impossible. Nevertheless, despite the absence of definitive knowledge of the environment of early Earth, certain principles of prebiotic chemical research have achieved general acceptance, providing a guide to the constraints of prebiotic reactions:[13] the substrates required for a synthesis must have been available in sufficient amounts at the site of the reaction; water, or anhydrous conditions should be used; and the yield should be significant enough that the products formed could be of further use to prebiotic chemistry.

Although geochemical evidence points towards a narrow geological window when life must have begun, there is little geological evidence from this period. Rocks on Earth are cycled, meaning that there are few rocks dating from the Archaen aeon, and many of those have had their fossil record destroyed by metamorphosis.[14] Fossil evidence dating back to 3.5 Gya proves that life – of approximately the complexity of simple bacteria today – was present at that time.[14,15] Today's simple bacteria are nevertheless a relatively complex life form, indicating that life must have started sometime before this point to be so relatively evolved. In fact, it was probably as evolutionarily complex as the last universal common ancestor (LUCA), which lived approximately 3.2–3.8

Gya[5] – and it is possible to reconstruct the genome of LUCA by alignment of overlapping gene clusters.

One study analysed gene clusters involved in genetic processing in a variety of bacteria and archaea. From this we are able to elucidate that LUCA had proteins with a functional role in the processing of genetic information (transcription, anti-termination control, translation and secretion) and therefore that LUCA had DNA, mRNA, tRNA and rRNA (which worked with the ribosomal proteins coded for by many of the genes), and two elongation factors.[16] From these studies, we are able to elucidate that LUCA must have had a cellular structure, which means it must have also had a mechanism of cellular division: LUCA was essentially as complex as a modern cell, but the evolution of LUCA itself cannot be assessed by phylogeny.[17] Generally, we assume that complexity arrives from something simpler: it seems that life did not begin with such complexity, but it is far from clear exactly how life did start and how it evolved to such a level of intricacy in a relatively short space of time.

Therefore, conceptual reduction to the simplest possible life form still provides us with a (at least, in the view of prebiotic researchers) intimidatingly complex cell formed of integrated genetic, catalytic and compartment-forming systems all linked by metabolism. Of course, it is easy to imagine how these may be simplified – today, for example, metabolism is tightly controlled by enzymatic catalysis and feedback mechanisms, but it was certainly initially simpler, less ordered and less selective – but even so it is hard to conceive how prebiotic chemistry could make these various subsystems. Surely different chemistries are required for each of these systems, and these chemistries would interfere with each other, descending into an inextricably complex mixture? Early experiments appeared to corroborate this position.

1.3 Approaches to understanding the origin of life

Experiments in prebiotic chemistry formally started with the famous Miller-Urey experiment in 1953. Electrical discharge (simulating lightning) was run through a strongly reducing mixture of gases (hydrogen, methane, ammonia and water) believed, at the time, to constitute the early Earth's atmosphere (***Figure 1.4***). Analysis of the products by paper chromatography revealed that a variety of amino acids **7** were formed: glycine **Gly**, and alanine **Ala** were undoubtedly present, with aspartic acid **Asp** possibly also present in smaller amounts;[18] it was later shown that this mixture could

support bacterial life.[19] Repetition of this experiment and analysis of the products using modern analytical techniques revealed that racemic mixtures of 17 amino acids were formed, and that asymmetric resolution of these could be obtained by crystallisation.[20]

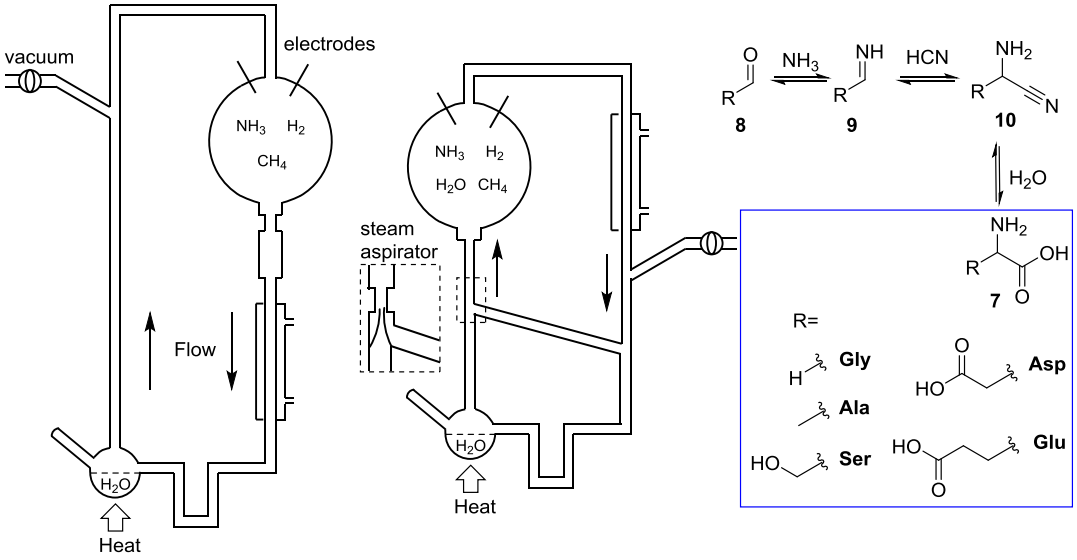


***Figure 1.4**: The experimental set-up used by Miller in his spark discharge synthesis of amino acids. The original set-up (left) and the modified version equipped with a steam aspirator to simulate a steam-rich volcanic eruption (centre) are both shown. Also shown (right) is a generalised scheme for the Strecker synthesis of amino acids, with selected amino acid side-chains shown in the blue box.*

Miller also carried out a variation of this experiment with intermittent addition of cyanamide **4** in 1958: these products were recently analysed, and fresh samples produced by repeating the experiment.[21] A wider variety of amino acids were produced: **Asp**, glutamic acid **Glu**, serine **Ser** and a number of unnatural amino acids, along with the simplest amino acid, **Gly**; glycine-containing dipeptides were also formed.[21] The products of another variation performed and analysed by Miller over 50 years ago,[22] with intermittent injection of steam and gas simulating spark discharge synthesis by lightening in a steam-rich volcanic eruption (experimental set-up shown in ***Figure 1.4***), have also been recently discovered and re-analysed.[23] 22 amino acids and five amines were identified – a higher diversity of amino acids than from the products of Miller's original set-up analysed at the same time – and hydroxylated compounds (such as **Ser**) preferentially formed. It is possible that hydroxyl radicals were formed from steam injection into the spark discharge that could have reacted with amino acid precursors or the amino acids themselves. Although Miller's highly reducing primordial atmosphere is now considered unlikely to be widespread on the early Earth,[23,24] and it is thought that

that Miller overestimated terrestrial energy dissipation by lightning,[25] Miller-style reactions could have been localised to reducing gases and lightening associated with volcanic eruptions,[23,26] with other energy sources such as UV predominating in other localised atmospheres.[27] It is of course also possible that amino acid synthesis in this manner could occur elsewhere in the solar system.[20]

Further experiments by Miller[28] revealed that the amino acids detected are formed from the Strecker synthesis (***Figure 1.4***, right), using aldehydes **8** and hydrogen cyanide formed *in situ*. Reaction of an aldehyde, **8**, with ammonia (present in the atmospheric model used in Miller's experiment) forms an imine **9**; nucleophilic attack of cyanide produces an α-aminonitrile **10.** Hydrolysis of aminonitriles produces the corresponding amino acids **7**. However, Miller noted that aldehydes, cyanide and ammonia appear predisposed to react: the conditions employed to generate the aminonitriles **10** in this reaction are not required for the generation of amino acids **7**.[28] The spark discharge experiment also produces the initial precursors (cyanide and formaldehyde **1**) required for nucleoside synthesis, but the nucleotides themselves are not observed. Without intrinsic chemical control, a complex mixture results, but one that does not reflect the complex interlinked systems required for life. It is not surprising that many prebiotic researchers promoted the reductionist theory that one particular system arose first and 'invented' the others (although without consensus as to which system), even if this required an implausibly simple mixture of substrates.

1.3.1 Metabolism-first

Gunter Wächterhäuser has been one of the main exponents of the idea that metabolism preceded genetics, and proponent of the iron-sulfur world theory. His hypothesis, set out in an extensive theoretical paper, was that life started with an autotrophic and autocatalytic metabolism in a two-dimensional mono-molecular organic layer anionically bound to positively charged mineral surfaces at an interface with hot water.[29] The thermodynamic equilibrium in surface metabolism favours the synthesis, not degradation, of polymers, with large polyanionic molecules (which have stronger surface-bonding abilities) selected. The earliest organism, according to this theory, consumed carbon monoxide and carbon dioxide, with its metabolism receiving reducing power from the reaction of iron sulfide and hydrogen sulfide to form molecular hydrogen and pyrite (iron disulphide); the pyrite can be used as a binding surface. Metabolism would be self-organised on the pyrite surface, initially consisting of the

reductive formation of methyl mercaptan and subsequent carbonylation to activated thioacetic acid (similar to the reductive acetyl-coenzyme-A pathway), which then fed into a carbon-fixation cycle (retrodictively constructed from the present-day reductive citric acid cycle).[30,31] This chemistry presumably could have occurred in deep sea vents, where superheated water containing dissolved hydrogen sulfide and transition metal sulphides suddenly mixed with excess cold sea water, causing the precipitation of sulfides (including iron sulfide).[13] Further papers were published, both theoretical with further information on how pyrite could induce chirality;[32] and reports of experiments that had been successfully carried out under Hadean conditions, proving that pyrite and molecular hydrogen can be formed from hydrogen sulphide and iron sulphide,[33] the formation of acetic acid from carbon monoxide,[30] the formation of amino acids,[34–36] and their condensation to form peptides.[37–39] It is hard to see how nucleotide synthesis is compatible with the conditions required for this iron-sulfur world, which indicates the need for an earlier genetic system.

Another metabolism-first theory set out in detail is the chemiosmotic origin of life.[40] The proponents of this theory note that LUCA, in addition to having a genetic system based on DNA and RNA and a catalytic system based on proteins, generated ATP *via* chemiosmotic coupling. Chemiosmosis involves the flow of ions across a membrane down an electrochemical gradient, but archaeal and bacterial membranes differ in their exact composition and in their biosynthesis suggesting that the membranes themselves were acquired after LUCA. This has led to the proposal that chemiosmosis is an intrinsic property of life, inherited from the time and location from which it arose: supposedly deep sea hydrothermal vents, which display a proton gradient across an inorganic membrane the same as that seen in an extant cell. Alkaline hydrothermal vents contain metal sulfide vesicular deposits, which could have been the site of a redox interface between the hydrogen-rich alkaline hydrothermal waters and a carbon-dioxide-rich acidic ocean. It is proposed that the electron transfer from molecular hydrogen to carbon dioxide across this inorganic membrane formed methane and acetate and released energy which could have been harnessed to fuel (proto-) biosynthetic pathways: thus life would have started as a by-product of carbon dioxide hydrogenation. This cell, with an inorganic membrane, would have evolved biosynthetic pathways to all the biomolecules required for life (it has been suggested that nucleotides would be easily formed under abiotic conditions in alkaline vents,[40] but this statement is not based in evidence) before ‘bubbling off’ the vent to escape, at which

point it would have developed an organic membrane. Although the reduction of carbon dioxide does take place in the presence of inorganic precipitates, the reaction proceeds only to equilibrium with no evidence of a proton gradient driving force.[41] The ability of the cells to acquire membrane-located proton pumps to maintain the pH gradient once they have escaped the hydrothermal vent has also been questioned.[42] It has even been suggested that the metal and metal sulfide clusters employed as co-factors in some enzymes were captured after the emergence of genes and gene-translation,[41] although it is more likely that metal ions would provide an anchor around which early functional peptides organise.[43]

Although metabolism-first theories have some advantages (circumventing the problem of how organic molecules were selected and continuously delivered to the site of life), they have also been robustly criticised.[41,42,44] In particular, the requirement for self-organisation of complex reaction systems using mineral and small metal catalysts that could both discriminate between similar substrates[44] but also be capable of catalysing multiple cycles,[45] and that the invention of a genetic system was a late addition to a complex evolved life form, stabilising a system with substantial information content,[44] are doubted. It should be noted, however, that although a metabolism-first origin of life in the manner described above is implausible, another theory of the origin of life (such as the RNA World) does not preclude some sort of proto-metabolism that generates the substrates required for the synthesis of the necessary biomolecules.[44–46] Indeed, some sort of reaction network is certainly required. Analysis of metabolism-related enzyme structure indicates that metabolic pathways evolve and concentrate around a core set of metabolites:[47] it is possible that these metabolites have been conserved from the earliest form of metabolism, even if the pathways between them have evolved.

1.3.2 Membranes

Some form of compartmentalisation at the origin of life is probable, as life is unlikely to begin in a very dilute solution.[48] A semipermeable membrane would confer a distinct chemical micro-environment on its contents and could protect oligonucleotides (preventing their random mixing or cross catalysis and therefore allowing Darwinian evolution to occur) whilst allowing small molecules to diffuse in and out.[49] A membrane would also allow molecules that are involved in each other's synthesis to stay together, which is surely advantageous for evolution.[13] It is hard to imagine that life could have started from membranes alone, suggesting a co-evolution with at least one other

system. Relatively simple mechanisms have been shown to enable competition for fatty acids between model protocells,[50] with feedback mechanisms mediating growth.[51] Accordingly, most theories for the origin of life include the presence of a membrane early on in evolution.

It is possible that membrane-formation and RNA polymerisation were closely linked at the origin of life. Nucleic acids can assemble on mineral surfaces, such as montmorillonite clay; montmorillonite can also increase the rate of vesicle formation (other silicates, such as aluminosilicate and borosilicate can also accelerate vesicle assembly).[49] The surface of montmorillonite clay is negatively charged, with a loosely bound layer of positively charged counter ions, which can attract negatively charged vesicles and free fatty acids.

It is not entirely clear, however, how the membranes of earliest cells survived standard biochemical processes. Non-enzymatic RNA replication (like enzymatic RNA replication) requires Mg^{2+} at a concentration (50–200 mM) sufficient to precipitate fatty acids as Mg^{2+} salts, thereby destroying the cell membrane.[52] Szostak and co-workers discovered that the presence of citrate can resolve this problem.[52] When Mg^{2+} is chelated by a dianion, such as citrate, it still has enough free co-ordination sites to catalyse RNA synthesis, but does not induce the precipitation of fatty acids or the degradation of single-stranded RNA. There is also a small (but reproducible) decrease in the melting temperature of double-stranded RNA in the presence of citrate, meaning the strands are more easily separated to allow template-directed replication.[52] A short acidic peptide (which potentially acts in the same way as citrate in these *in vitro* experiments) is now present in the active site of RNA polymerase, indicating a possible role for short peptide co-factors at the origin of life.[52] It has also been proposed that early life forms could have used other, more stable, membrane components. For example, a fatty acid glycerol ester and its amide analogue (which can be synthesised *in situ* by an enzyme entrapped within the protocell) are capable of stabilising oleic acid vesicles in the presence of Mg^{2+}.[53]

1.3.3 Biopolymers

Each biopolymer has a particular set of monomers from which it is polymerised with a single type of consistent chemistry; this allows existing polymer folds to adapt to new functions through mutation and recombination.[54] The first polymers would have linked

the world of abiotic chemistry and the biotic world of biocatalysis, and so should be formed from monomers that can be easily synthesised by abiotic chemistry, able to facilitate their own replication, and able to promote a wide range of chemistry.[54] Unfortunately, no biopolymer has been proved to fulfil all these conditions: although RNA can achieve the latter two conditions, the difficulty in forming nucleotides and their initial abiotic polymerisation has led some to suggest that a different genetic mechanism preceded the one we know today.

1.3.4 Peptides

Nucleic acids (and their possible precursors) are more popular candidates for the first biopolymer than proteins: nucleic acids have a structure that intrinsically allows information storage and transfer as well as displaying catalytic abilities; proteins have less capacity for information storage and therefore are likely to have played only a phenotypic role at the origin of life. Nevertheless, a protein interaction world has been proposed, involving life dependant on a self-reproducing and expanding system of protein interactions, although experimental evidence for this is absent.[55] Replication in this theory is an abstract communication system based on the replication of interactions between molecules, rather than the replication of the molecules themselves: RNA would later emerge to reproduce these peptide interactions.

It has long been emphasised that the popularity of nucleic acid-based theories of early life does not preclude the presence of amino acids on the early Earth.[56] Although it is perhaps the most-well known, the Strecker reaction is not the only prebiotically plausible synthesis of amino acids. The Bucherer-Bergs hydrolysis of amino nitriles (which could have been formed *via* the Strecker reaction) shown in ***Figure 1.5*** (blue arrows) involves addition of bicarbonate and carbon dioxide to the amine, resulting in the formation of a hydantoin **13**, a precursor to amino acids **7** *via* a two-step hydrolysis.[57,58] A similar directed hydration occurs when simple carbohydrates are used in place of carbon dioxide (***Figure 1.5***, red arrows);[59–62] the use of chiral sugars mediates the formation of enantioenriched aminoamides.[63] The presumed greater concentration of carbon dioxide over simple carbonyls in the early atmosphere, however, has led to the suggestion that Bucherer-Bergs synthesis would have taken place preferentially over carbonyl-mediated hydrolysis.[61]

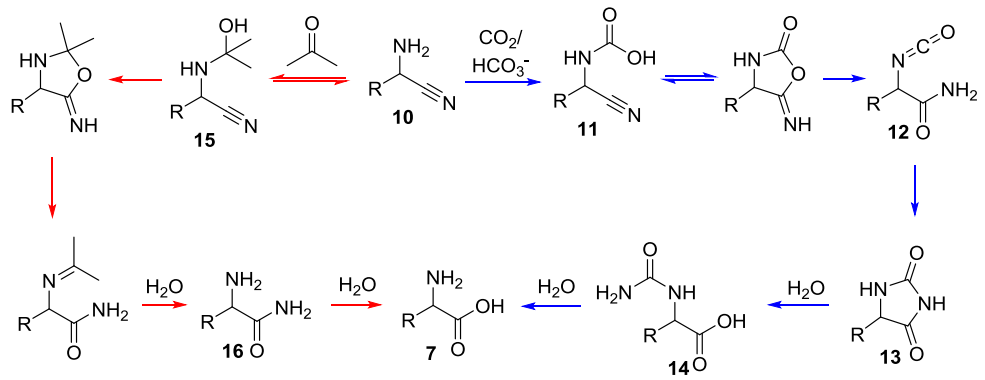


***Figure 1.5**: Methods for the directed hydration of amino nitriles **10**. The Bucherer-Bergs hydrolysis (blue arrows) uses carbon dioxide/bicarbonate: a carbamic acid **11** is formed with a carboxylated well-placed to attack the tethered nitrile. The first cyclic intermediate collapses to give an isocyanate **12**, recyclises to form a hydantoin **13** and undergoes two-step hydrolysis via N-carbamoylamino acid **14** to amino acid **7**.The use of a carbonyl compound (such as acetone) as a promoter results in aminal **15**-promoted hydrolysis to the corresponding amide **16** and subsequently the amino acid **7**.*

Retrosynthetic analysis of the Strecker reaction allows identification of the aldehydes required to access the canonical amino acids. More recently, the elucidation of a cyanosulfidic network that uses Kiliani-Fishcher-type homologation of hydrogen cyanide with Cu^{I}-Cu^{II} photoredox chemistry has shown how amino acid precursors, some of which are also common to nucleotides and lipids, can be formed under a geochemical scenario now thought to be more prebiotically plausible than the one used by Miller.[64]

Given that amino acids were plausibly available on the prebiotic Earth, it is possible that they, and their oligomers, could have supported the advent of life. Further support is lent by the catalytic and functional abilities of peptides. Many biological reactions are catalysed by (protein-based) enzymes. The ribosome – the translational machinery of the cell – is built around interactions between RNA, amino acids and proteins,[65] indicating a common evolution of these species.[66,67] Indeed, it has even been shown that peptides derived from the ribosomal core enhance the RNA polymerase activity of a ribozyme.[68] Early RNA self-replication could therefore have been dependent on the availability of short peptides:[68] these could have reduced RNA dependence on metal cations such as Mg^{2+}, which, at the high concentration required, is incompatible with membranes.[52] Peptides have also been shown to protect fatty acid membranes from

high concentrations of Mg^{2+} (and single stranded RNA from Mg^{2+}-mediated degradation) required for non-enzymatic templated RNA polymerisation.[52]

Prebiotic peptides were not necessarily associated with RNA: the smallest natural proteins capable of forming a stable structure without interacting with other biomolecules are just 20 amino acids long.[43] It is possible that even shorter peptides could act as catalysts on the early Earth, especially in conjunction with a positively charged metal ion, which would provide an anchor around which the peptide could organise (and stabilise), as well as being useful for catalysis.[43] Albeit inefficient and non-specific functional peptides can be very small: a dipeptide (itself produced *in situ* by another dipeptide) has been shown capable of enhancing vesicle growth by associating with the vesicle membrane and recruiting fatty acids from neighbouring vesicles.[50]

Longer peptides are capable of auto-catalytic synthesis: an α-helical peptide, 32 amino acids long, accelerates the formation of a thioester-promoted amide bond between its precursor 15- and 17-residue fragments. [69] (It is therefore also self-replicating, as long as there is a supply of the precursor fragments.) The same peptide, when in the presence of another self-replicating peptide with which it shares a precursor fragment and a supply of their substrate fragments, displays interesting catalytic abilities. Ordinarily, one would expect one of the self-replicating peptides to outcompete the other; instead, they catalyse each other's production.[70] This type of system could provide selection based on feedback processes of genotype replication in an early life form.[70]

The formation of peptide bonds in water, however, is thermodynamically disfavoured (as water produced in the condensation reaction must be removed).[57,71] It therefore follows that once the α-amino acids are formed, they must be activated to allow polymerisation to give peptides and (eventually) proteins.[72] Prebiotic chemical explorations of peptide formation have generally started from the preformed monomeric amino acid, based on the reasonable assumption that the shortest and most direct reaction pathways are the most plausible, and in the knowledge that amino acids can be produced under prebiotically plausible conditions.

A suitable activation of amino acids to allow their polymerisation, however, is still not clear. Several strategies involve the intermediary of cyclic structures which are activated towards polymerisation, such as *N*-carboxyanhydrides **17**[73] (which can form from *N*-carbamoylamino acids **14**, an intermediate in the Bucherer-Bergs synthesis of amino acids[57,58]) and acylphosphoramidates **18** (***Figure 1.6****, blue box*),[74–76] but these are inefficient and have not been shown to oligomerise a wide range of amino acids. Peptide bond formation promoted by α-hydroxy acids has also been investigated,[77] involving the formation of oligomers containing a combination of amide and ester linkages (but driven by harsh wet-cool/dry-hot cycles) followed by ester-amide bond exchange and ester hydrolysis (***Figure 1.6***).

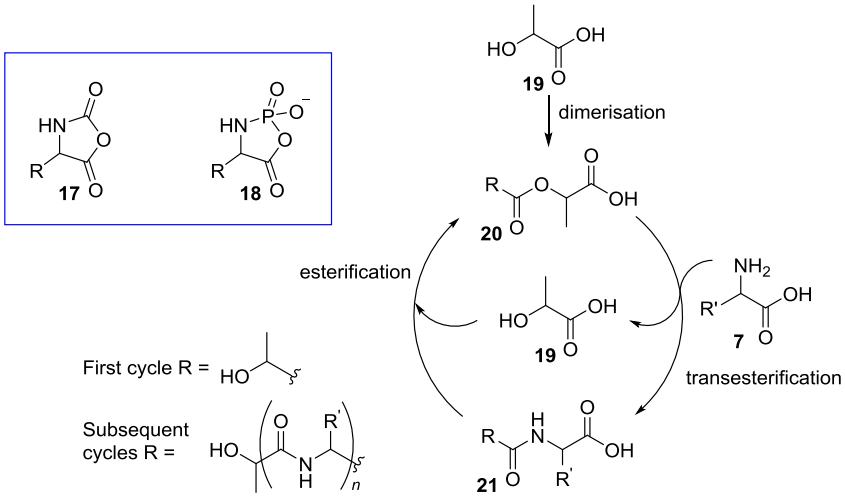


Figure 1.6*: Selected activation strategies of amino acids. Cyclic intermediates such as N-carboxyanhydrides* ***17*** *and acylphosphoramidates* ***18*** *(blue box) are activated towards polymerisation. A proposed reaction scheme for lactic acid* ***19****-mediated peptide bond formation and depsipeptide (a peptide containing ester linkages, ie* ***20*** *in subsequent cycles) elongation by ester-amide exchange is also shown.*

1.4 The RNA World

The discovery of ribozymes,[78–80] which confirmed earlier speculations that the first life forms could have depended on just one biopolymer with RNA acting as both an informational and functional molecule,[81,82] led Walter Gilbert to propose the term 'RNA world theory'.[83] Different researchers often mean different things by 'RNA World hypothesis'; however, all agree with Joyce's simple definition that it is a point in time at the origin of life when: genetic continuity was assured by the replication of RNA; Watson-Crick base-pairing was the key to this replication; and genetically encoded proteins didn't act as catalysts.[84]

The flow of information through the earliest forms of RNA-based life would therefore have been significantly simplified compared to today (***Figure 1.7***). The RNA world offers perhaps the most parsimonious origin of life – it is far more likely that one molecule was synthesised by random chemical reactions than two different, but co-dependant, molecules were synthesised in the same place at the same time[85] – but the theory does not preclude the coevolution of other biomolecules alongside RNA. The RNA world theory also offers continuity: it is not clear how RNA could replace a self-replicating system based on a different (and unknown) system.

Figure 1.7*: The central dogma of molecular biology plausibly adapted for the RNA world. The same RNA molecule could act as both genotype and phenotype, but it is perhaps more likely that one RNA molecule acted as a store of genetic information, and when replicated, its daughter folded in a specific conformation that contributed to its function.*

All organisms today have a ribosome in which RNA plays a key role, implying that RNA also had this role in LUCA. Specifically, we know that LUCA had ribosomal RNA (and therefore that RNA was acting as a critical functional molecule in the synthesis of proteins) and that it had a wide variety of RNA co-factors implying a level of metabolic complexity.[86] The number and complexity of enzymes responsible for the synthesis of RNA nucleotides, and the fact that DNA biosynthesis is a derivative of RNA biosynthesis, strongly suggests that RNA preceded DNA.[87] Phylogenomic analysis indicates that most extant enzyme activities are associated with just nine ancient and wide-spread protein structures involved in nucleotide metabolism.[88] This implies that the first enzymes took over the synthesis of nucleotides, which had previously been produced by an RNA-based biochemistry in an RNA world. The prevalence of co-factors based on nucleotides and nucleotide fragments today can be viewed as a 'molecular fossil' of metabolism in the RNA world.[89]

Many different theories abound, however, as to the nature (if any) of life preceding the RNA world, the metabolic complexity of the RNA world and the role co-factors (both metallo and bioorganic) played. An origin of life which depended on RNA is sometimes referred to as the 'strong', 'purist' or 'reductionist' RNA world hypothesis.[90,91] Others believe that a hybrid world, with a loosened interpretation of the theory and the

significance of RNA, is more likely.[2,13,92] All living organisms have similar metabolic networks, implying that all present forms of metabolism descended from the same network.[48] Some metabolic reactions can be catalysed by metal ions present in Archaean sediments; RNA and short peptides in this context could act as only a support for the catalyst, increasing metal solubility and local concentration, and increasing the reaction specificity.[48] Others imagine a more prominent role for RNA: inorganic catalysts assisting the synthesis of small and simple RNA molecules, which, due to their greater catalytic efficiency, become ever more complex and dominant.[93]

Eschenmoser recognised that life and its origin are immensely complex; the origin of a molecular structure upon which life depends and which allows life to occur is still complex, but perhaps simple enough to study in the laboratory.[94] The plausibility of the RNA world hypothesis, and the importance of RNA on the early Earth, is theoretically examinable by synthetic chemistry. If RNA can be synthesised under prebiotic conditions in a short time-scale in a laboratory, it is also possible that it can be synthesised over an aeon on prebiotic Earth. If other molecules associated with life can be synthesised under the same conditions, or even with interlinking syntheses, support for the 'loosened' RNA world hypothesis is strengthened. Of course, there would have been several stages associated with the synthesis of RNA on the early Earth:[13] the nucleotides must first be formed, then polymerised to give oligonucleotides of random sequences, then copied or replicated. Once a pool of RNA sequences has formed, natural selection can place. The latter can be investigated by studying RNA evolution; the first three stages can be investigated by prebiotic chemistry researchers, although the role of chemical evolution should not be discounted.

1.4.1 The contemporary RNA World

The ubiquity of RNA in life today offers a hint as to how it could have dominated the earliest forms of life. RNA today plays a critical role in life, with RNA molecules carrying out a range of functions: some purely RNA centered, and others relying on proteins and DNA to complete their role.[85] Many co-factors are derived from RNA nucleotides (***Figure 1.8***).[89] RNA itself plays a central role as a messenger (mRNA) and translator (tRNA) in gene expression;[85] it is present as a biocatalyst in the ribosome (rRNA) and telomerase (a ribonucleoprotein which ensures all genetic information is transmitted with each round of replication and cell division);[85,95,96] it regulates DNA gene expression;[85] and binds specific small metabolites and uses the binding energy to

switch to a different RNA structure (riboswitch).[85,97] Once the genetic information stored as DNA has been transcribed into mRNA and undergone initial processing in the cell nucleus (to remove some of the non-coding sections that are not translated into a functional protein), it is transported to the ribosome where the genetic sequence of the mRNA is translated into proteins. Transfer RNAs (tRNAs) carry amino acids to the ribosome and interact with mRNA to ensure the amino acids are added to the growing peptide in the correct sequence.[65,98]

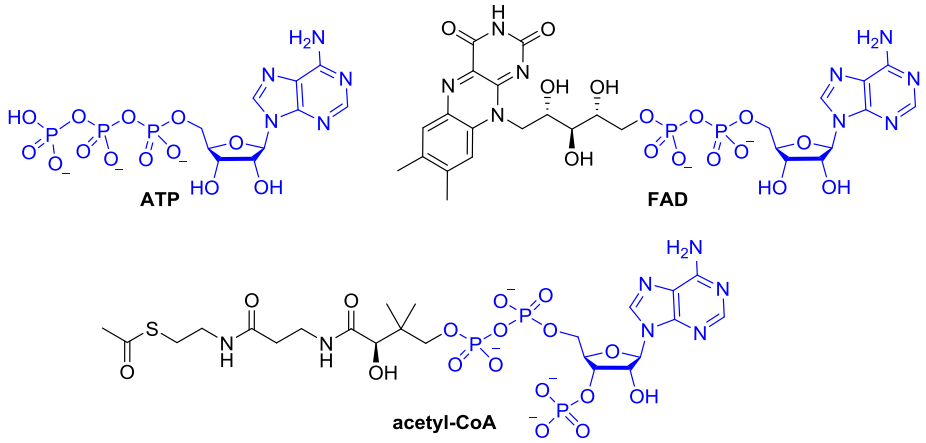


***Figure 1.8**: Three adenosine-based co-factors: adenosine triphosphate (ATP), flavin adenine dinucleotide (FAD) and acetyl co-enzyme A (acetyl-CoA). The ribonucleotide portions are highlighted in blue.*

The discovery of ribozymes[78–80] (a ribonucleic acid with catalytic activity; the name is a portmanteau of 'ribonucleic acid' and 'enzymes') demonstrated the full functional potential of RNA.[85] The first ribozyme to be discovered was a self-splicing RNA: it undergoes transcription in the same way as most other RNAs, but remarkably doesn't require enzymatic processing before translation.[79] Instead, it has the the intrinsic ability to perform the three cleavage and two ligation reactions involved in splicing, implying that enzymes are not involved in these reactions *in vivo*.[79] This ability to self-splice may have played a critical role in evolution: it would allow genetic recombination – a role fulfilled by sex today.[83] This ribozyme, like most other ribozymes, therefore carries out phosphoryl transfer reactions (cleavage and ligation of the sugar phosphate backbone).[97,99] Phosphoryl transfer reactions can be catalysed in a number of ways by ribozymes: they can act as a general acid/base catalyst, stabilise charge build-up, or simply position the substrates in the optimum conformation for reaction.[97,99] RNA is capable of forming specific, high-affinity pockets for divalent metal ions, which can also play a catalytic role.[97]

The self-splicing RNA, once processed, is an rRNA: destined to fulfil a functional role in the ribosome. The ribosome mediates interactions between mRNA and tRNA and therefore determines the sequence of proteins formed,[65] and also catalyses peptide bond formation.[95,100] Peptidyl transfer is catalysed by RNA not proteins, although the exact mechanistic details are unknown: rRNA could certainly act as an acid/base catalyst or stabilise charge build-up as it does in phosphoryl transfer reactions.[97] Mutations at the putative catalytic residue can be tolerated, leading to suggestions that peptide bond formation occurs as a result of substrate positioning rather than chemical catalysis,[101] but the 2′-hydroxyl moiety of this nucleotide is believed to play a catalytic role.[99]

The abundance of RNA in the ribosome reflects its potential origin as an entirely RNA structure, with proteins mainly at the surface, perhaps added later to enhance structure and function.[65] Furthermore, a greater proportion of protein within a specific functional site of the ribosome reflects its antiquity: the E site (where the spent tRNA exits the ribosome) is the most recent addition and is protein rich; the decoding centre is older and is mostly RNA; whilst no proteins come within 18 Å of the peptidyl transferase centre, demonstrating (if this argument holds true) its genesis as the only functional site of the ribosome.[65,95] Although rRNA is associated, and works, with protein, many of the important roles of the ribosome involve and depend on direct RNA–RNA interactions, such as the peptidyl transferase centre, mRNA start-site selection, codon-anticodon interaction and decoding.[85]

It is not clear whether first ribosome developed simply to catalyse peptide bond formation, or whether it evolved from a structure with a different function: it has been proposed that, before being adapted for protein synthesis, the ribosome originally performed essential tasks in the RNA world,[98] possibly being used as an organising centre to correctly orientate nucleotide-chaperoned metabolites to promote reactions.[102]

The functional dominance of proteins in today's life makes it hard to retrodictively analyse which reactions essential to life could have been catalysed by RNA in the past. From *in vitro* evolution of a pool of random RNA sequences, however, we can understand the range of reactions that RNA is capable of catalysing. Many studies have elucidated the plausibility of an RNA-based metabolism: RNAs are capable of

catalysing RNA polymerisation,[103–105] the assembly of nucleotides,[106] the activation of amino acids,[107] the charging of tRNAs with amino acids,[108] and Diels-Alder cycloadditions.[109] The early ribozymes could be very small: remarkably, a pentanucleotide sequence, just long enough to fold to create an active site, has been shown to be capable of aminoacyl transfer.[110]

1.4.2 Oligomerisation

RNA-catalysed RNA replication provides a chemical basis for natural selection, and therefore Darwinian evolution. Before evolution arose it is unlikely that a self-replicating ribozyme could arise; but, conversely, without self-replication it is impossible to perform an evolutionary search for the first self-replicating ribozyme. It is therefore most likely that the RNA oligonucleotides, themselves formed from random polymerisations of monomers, were replicated by non-enzymatic template directed synthesis.[84] The non-enzymatic oligomerisation and copying of RNA, however, is seen as one of the major problems facing those investigating the RNA world.[13]

Replication by non-enzymatic template directed synthesis requires the presence of an RNA molecule to be replicated – which must have been formed from non-enzymatic non-templated directed random oligomerisation. Nucleotides contain three dominant nucleophilic groups: the 3′-OH, the 2′-OH, and the 5′-OH, in order of increasing reactivity. Therefore, the reaction of several nucleotides will lead to an oligonucleotide containing a number of 2′,5′-phosphodiester linkages (***Figure 1.9***). An oligonucleotide consisting only of ‘natural’ 3′,5′-linkages will only be produced in the presence of a catalyst (such as RNA polymerase in our cells) that can absolutely control the regiochemical outcome of phosphodiester ligation. Furthermore, oligomerisation does not occur spontaneously in aqueous solution. Investigations using external activating agents (such as cyanamide **4**) have generally been unsuccessful, yielding only very short oligomers in poor yields; the polymerisation of preactivated nucleotides has proven more successful.[13] Nevertheless, oligomerisation of adenosine 2′,3′-cyclic phosphate has been show to occur in the dry state with prebiotically plausible general acid-base catalysts such as amines and imidazoles[111] with hexamers forming after 40 days,[112] although these oligomers contain a 1:2 mixture of 2′,5′- and 3′,5′-linkages and terminated with both 2′- and 3′-monophosphates.[111] Non-enzymatic template-directed synthesis has been investigated rather more thoroughly than non-templated

polymerisation, however, and so nucleotide polymerisation will be discussed in reference to templated synthesis.

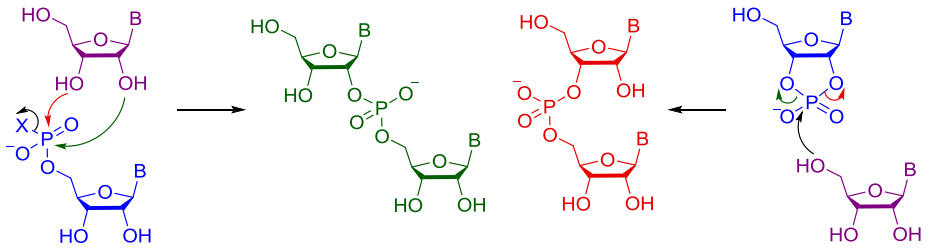


***Figure 1.9**: Attack of a nucleoside (purple) on a nucleotide (blue), activated by the presence of a 2',3'-cyclic phosphate (right) or a leaving group on the 5'-phosphate (left), can lead to the formation of a 2',5'-linkage (green) or a 3',5'-linkage (red). B=nucleobase, X=leaving group.*

The presence of a template makes RNA replication slightly easier: complementary binding of a monomer to the template adjacent to the 3′-terminal nucleotide of the primer allows a means of aligning the 5′-phosphate of the monomer to nucleophilic attack by the 2′- or 3′-hydroxyl of the primer's 3′-terminus, which cannot occur in non-templated oligomerisation. In addition to the efficiency of oligomerisation (the rate at which monomers are incorporated into the growing chain) and regiospecificity (the formation of natural 3′,5′-linkages instead of unnatural 2′,5′ ones – non-enzymatic templated 3′-5′ oligomerisation can selectively take place, albeit very slowly[113]), fidelity (the inclusion of a complementary over a non-complementary nucleobase) must also be taken into account.[114] Unusual and potentially unproductive Watson-Crick base pairs can form,[115] but this causes oligomerisation to stall,[116] meaning that accurate copies are completed more rapidly and demonstrating a means for maintaining genetic fidelity on the primordial Earth.

Shorter nucleic acids are replicated faster and therefore outcompete longer ones, leading to ever shorter sequences and the loss of genetic information, but this may be partially solved by the local environment: temperature gradients can mitigate against dilution and degradation, and drive the replication of longer oligonucleotides.[117] The formation of very long oligonucleotides may not be required for function on the early Earth, as it has been shown that a pentanucleotide displays catalytic acitivity.[110]

Unfortunately, the oligomerisation of simple nucleotides (3′- or 5′-monophosphates) is energetically unfavourable to the point of near-impossibility on a significant scale. There are two obvious methods of prebiotic nucleotide activation: the addition of a

leaving group to the 5′-phosphate; or by using the ring strain intrinsic to a 2′,3′-cyclic nucleotide (***Figure 1.9***). Although nucleotide triphosphates (with a pyrophosphate leaving group) are used in biology today as activated monomers for polymerisation, they are also too unreactive for non-enzymatic polymerisation.[113] It is of course possible that extant polymerisations are novel cellular inventions, rather than prebiotic mimics.[118] Activation by 2′,3′-cyclic phosphate ring strain (still kinetically stable enough to prevent rapid hydrolysis in water[119]) could incur other advantages on the early Earth. Oligomer hydrolysis occurs *via* intramolecular phosphodiester transfer, forming a 2′,3′-cyclic phosphate, and offering a simple chemical strategy for re-ligation. Although 2′,3′-cyclic phosphates are hydrolysed to give a mixture of unactivated 2′- and 3′-monophosphate isomers, these can be cyclised to the original product using cyanoacetylene **5** (which itself is a substrate for the synthesis of 2′,3′-cyclic nucleotide monophosphates).[120] However, their direct polymerisation is both thermodynamically unfavourable and kinetically slow as the ring is only slightly activated towards nucleophilic ring-opening by a hydroxyl group.[121]

This inefficiency of oligomerisation implies that some other method of activation is required, but even attempts to copy RNA sequences using chemically activated ribonucleotides have been generally unsuccessful because monomer addition is inefficient and variable depending on the identity of the monomer.[122] Nevertheless, some approaches have been successful; several require the presence of imidazoles, which are thought to have played an important role in primordial chemistry.[123] One proposed method for the *in situ* activation and templated oligomerisation of nucleotides is through the use of an alkylated imidazole and carbodiimide (a tautomer of the prebiotically plausible cyanamide **4**), but this method has only been shown to work with the non-prebiotically plausible EDC **22**.[124]

Nucleotides activated with 2-methyl imidazole **23** are known to spontaneously polymerise when guided by a template sequence,[125] although different bases are incorporated with varying efficiencies.[126] Although both 2′- and 3′- hydroxyls are able to attack the phosphate of the incoming nucleotide, 3′,5′-linkages are predominantly formed when the incoming nucleotide is activated by **23**, thought to be due to optimal placement of the 3′-hydroxyl of the primer for in-line attack of the 5′-phosphate on the activated monomer.[127] The rate and fidelity of the reaction can be improved by using sulfur derivatives – which are prebiotically plausible[128] – of the canonical

nucleotides:[129–131] the resulting base pair is more thermodynamically stable than the canonical base pair. It is also known that the reaction is catalysed by the presence of an activated downstream monomer[132] or oligomer (***Figure 1.10***).[133,134] The activated nucleotide adjacent to the primer forms a covalently-bound imidazolium-bridged di- or oligonucleotide intermediate **24** which can form more Watson-Crick base-pairs with, and therefore bind more tightly to, the template than a monomer.[132,135] Attack of the 3′-hydroxyl of the primer on the imidazolium-bridged intermediate results in displacement of the activated nucleotide as the leaving group, rather than attack of the 3′-hydroxyl on the monomer with displacement of the 2-methyl imidazole **23** leaving group.

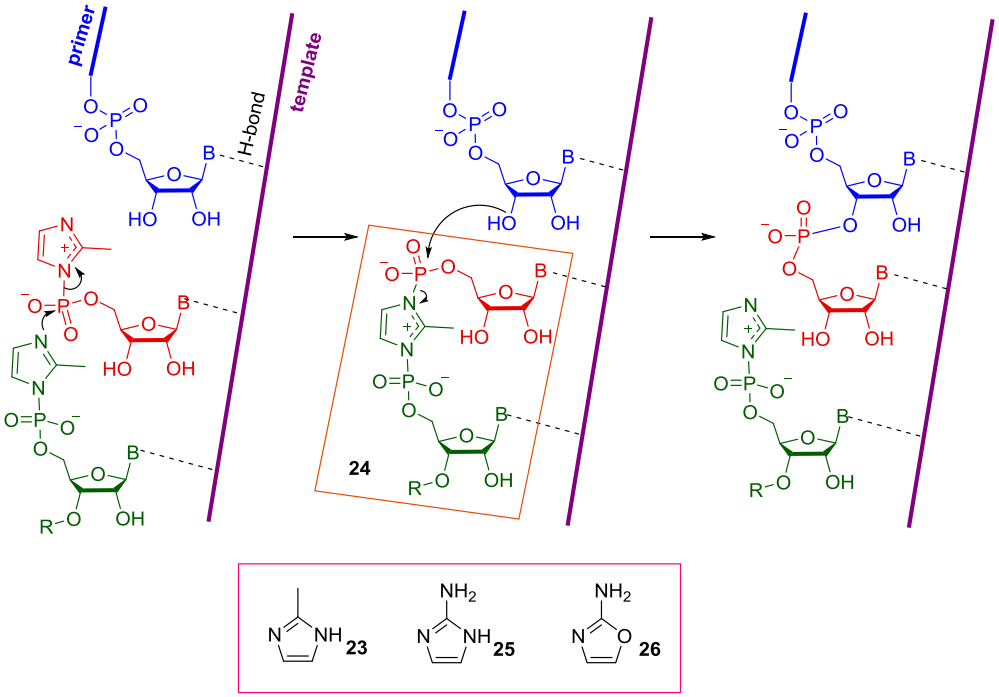


Figure 1.10*: Primer (blue) extension using nucleotides activated with 2-methylimidazole* ***23*** *hydrogen-bonded to a template sequence (purple). The activated nucleotide adjacent to the primer (red) reacts with the downstream monomer or oligomer (green), forming an imidazolium-bridged intermediate (****24****, orange box). Attack of the 3′-hydroxyl of the primer on* ***24*** *results in the formation of a 3′,5′-linkage with displacement of an activated nucleotide. The structures of* ***23****, 2-aminoimidazole* ***25*** *and 2-aminooxazole* ***26*** *are also shown in the pink box.*

2-aminoimidazole **25** has been shown to be an even more efficient activating agent than **23** (with a 6–100 fold rate advancement compared to **23**, depending on the monomer incorporated)[136] with the prebiotic advantage of being constitutionally similar to, and formed under the same conditions as, 2-aminooxazole **26**,[137] a key intermediate in the synthesis of pyrimidine nucleotides (discussed in detail later in this chapter).[138]

It is generally agreed that the ligation of short, templated oligonucleotides should offer a more efficient prebiotically-plausible route to generating longer RNA chains than solely by polymerisation of RNA mononucleotides.[54,122,139] Strand ligation will be favoured over hydrolysis by template-induced proximity, and template-chelation should also improve the thermodynamic favourability. Oligomers formed from the non-templated polymerisation of adenosine 2′,3′-cyclic phosphate, with both 2′- and 3′-monophosphate termini, are suitable substrates for ligation. When treated with sodium thioacetate **27** and an electrophile such as cyanoacetylene **5,** the free 2′-hydroxyls are selectively acetylated (although an explanation for this selectivity is unclear) *via* an acyl-transfer cascade involving a mixed carboxy-phosphate anhydride intermediate **28** (***Figure 1.11***).[139] This selectivity means that subsequent electrophilic activation of a 2′-phosphate results in rapid formation of a 2′,3′-cyclic phosphate (by attack of the free 3′-hydroxyl); the protected 2′-hydroxyl, in contrast, cannot generate a 2′,3′-cyclic phosphate, allowing ligation with the 5′-hydroxyl of the neighbouring oligomer to occur.[139] The templated ligation of oligonucleotides with 2′,3′-cyclic phosphate termini gives mostly 2′,5′-linkages, albeit slowly,[139] but these preferentially hydrolyse to 2′,3′-cyclic phosphates, which themselves preferentially hydrolyse to give 3′-phosphates, offering a further enrichment of canonical 3′,5′-linkages.[140]

RNA containing 2′,5′-linkages in addition to the natural linkages has an altered structure: the phosphodiester backbone is compressed and kinked.[141] The strict relationship between structure and function in proteins implies that a similar dependence on structure is required for efficient functionality in ribozymes, and therefore that RNA produced by methods that do not selectively form natural linkages has a sufficiently malformed structure to destroy functionality. 2′,5′-Linkages do not cause many problems, however: unnatural linkages do not prevent replication, for example, although they reduce its rate;[116,142] and RNA oligomers that can specifically bind small signalling molecules are still able to do so despite varying backbone heterogeneity.[142] It also appears catalytic function is not adversely affected: for example, the hammerhead ribozyme can still self-cleave at the correct specific site with varying backbone heterogeneity, although a higher proportion of 2′-5′ linkages slows the rate of self-cleavage and results in a lower maximal extent of self-cleavage (when containing 10% 2′,5′-linkages its activity reduced by 20%).[142] The presence of 2′-5′ linkages at two specific points in the ribozyme's active site resulted in substantial loss

of activity:[142] it is likely that the ribozyme evolved to exploit the C2′-OH of these points as a nucleophile or general acid.

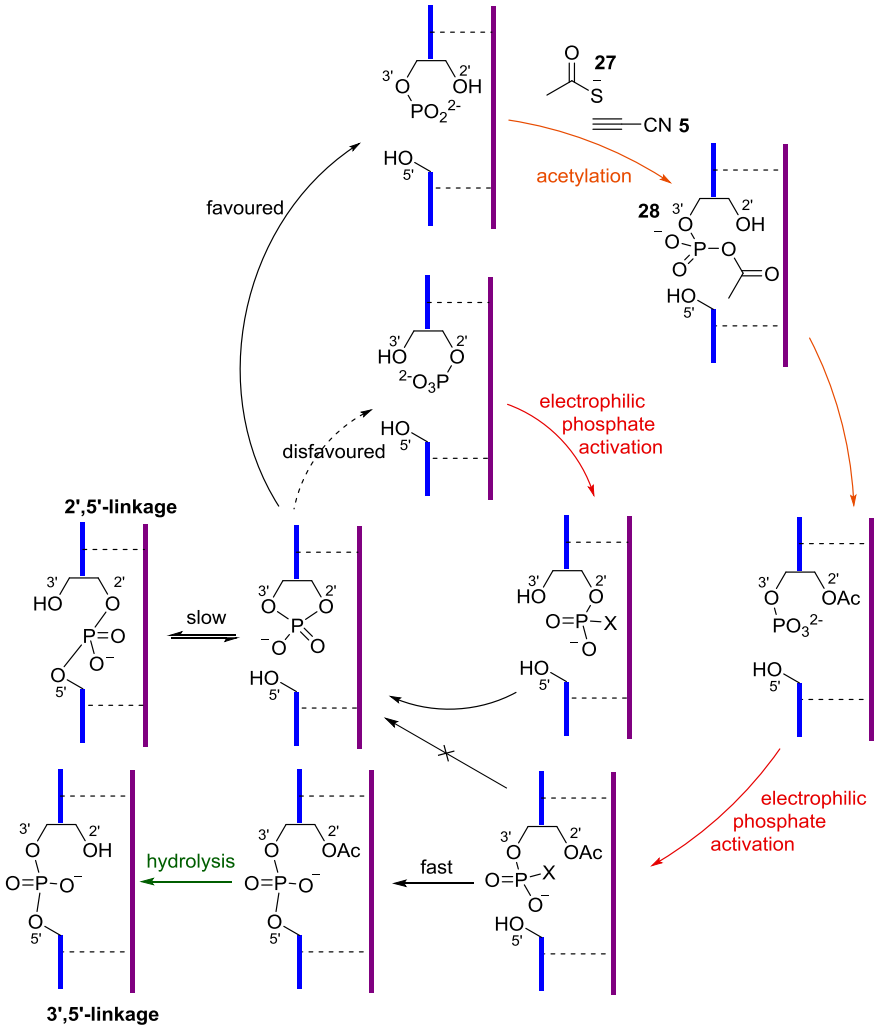


***Figure 1.11**: Chemoselective acetylation of 3′-phosphate terminal oligomers over the corresponding 2′-terminal oligomers in the presence of thioacetate **27** and cyanoacetylene **5** protects them against cyclisation to 2′,3′-cyclic phosphate termini upon electrophilic phosphate activation.3′-Phosphate terminal oligomers therefore readily undergo ligation reactions to generate 3′,5′-linkages; the unprotected 2′-phosphate termini form 2′,3′-cyclic phosphates which can undergo hydrolysis to 2′-phosphates (disfavoured) or 3′-phosphates (favoured), or slow ligation to give 2′,3′-linkages, which subsequently hydrolyse to reform 2′,3′-cyclic phosphates.*

The accumulation of structural variety resulting from the inclusion of 2′,5′-linkages decreases the overall stability of the RNA duplex. Base-pairing hydrogen-bonding is weakened (by an increase in buckle amplitude) and base stacking is destabilised (by varied and disadvantageous rise distances; the diminished base stacking in duplexes containing 2′,5′-linkages is demonstrated by the increase in UV absorbance compared

to duplexes containing all natural linkages upon heating[142]). Together, these reduce the enthalpy of duplex formation and therefore lower the melting temperature of the duplex[141] making templated oligomerisation and ligation possible. Backbone heterogeneity therefore allows strand separation without the loss of information transfer: this combination of properties may have meant that the presence of unnatural 2′-5′ linkages were not only tolerated, but perhaps even essential at the origin of life. A small degree of infidelity (and therefore sequence variation) would also be useful, generating selective advantage and therefore contributing to evolution.[143]

In addition, the inclusion of 2′,5′-linkages may have allowed early separation of genotype from phenotype whilst life was still RNA-based: ribozymes with a homogeneous backbone would have well-folded structures more suited to catalysis, whilst a ribozyme with an identical sequence of bases but a heterogeneous backbone would be more easily used as a replication template.[142]

1.4.3 Synthesis of nucleotides

Of course, the oligomerisation of nucleotides requires that nucleotides be present, presumably synthesised abiotically in the local environment. Elucidating a prebiotically plausible synthesis is another major problem that has yet to be solved. Nucleic acids are traditionally retrosynthetically disconnected to deoxy- and ribonucleotides; and these nucleotides are retrosynthetically disconnected to a sugar, a nucleobase and phosphate[144] (exemplified by the retrosynthetic disconnection of cytidine-5′-monophosphate ß-**29**-5′P in ***Figure 1.12***). It is therefore unsurprising that the prebiotic accessibility of these component parts as been perceived as an indicator of the prebiotic plausibility of nucleotide abiogenesis. The prebiotic availability and synthesis of ribose *ribo*-**30**, the canonical nucleobases, and phosphate has been studied with varying success.

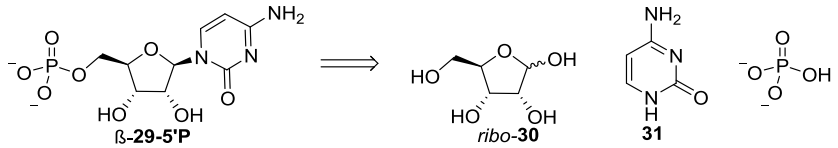


Figure 1.12*: Traditional retrosynthesis of cytidine-5′-monophosphate ß-**29**-5′P to ribose ribo-**30**, cytosine **31** and phosphate.*

1.4.3.1 Prebiotic synthesis of ribose

It has generally been assumed that ribose *ribo*-**30** required for nucleotide synthesis on the early Earth would have been generated through the formose reaction, or a variation thereof.[57] The polymerisation of formaldehyde **1** in basic solution, forming a complex mixture of sugars, was discovered by Butlerow in 1861,[145] with a plausible mechanism proposed a century later.[146] An initial period of slow reactivity is followed by a rapid autocatalytic process, which produces a complex mixture of sugars. The initiation step is slow because it is believed to require a homoaldol reaction of formaldehyde **1** (requiring an infrequently occurring umpolung mechanism so formaldehyde can act as a nucleophile), producing glycolaldehyde **32**, which then acts as an efficient catalyst (addition of **32** vastly increases the initial rate of reaction).[146] An oversimplified representation of the autocatalytic reaction cycle of the formose reaction is shown in ***Figure 1.13***. Forward and reverse aldol reactions and aldehyde-ketone tautomerisations allow the production of two molecules of glycolaldehyde **32** from each initial molecule of **32** and two of formaldehyde **1**.

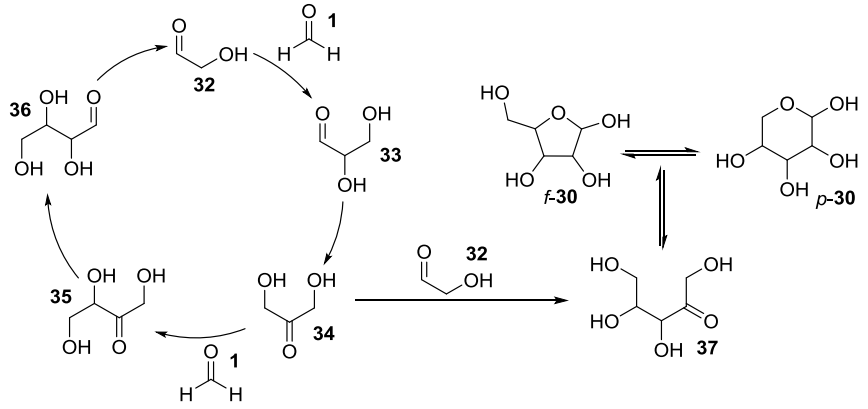


Figure 1.13*: Simplified autocatalytic reaction cycle of the formose reaction: two molecules of* ***32*** *are generated from every one* ***32*** *and two* ***1****. Reaction of* ***32*** *with dihydroxyacetone* ***34*** *generates ketopentose* ***37*** *which isomerises to the corresponding aldopentoses* ***30****.*

Glycolaldehyde **32** reacts with more formaldehyde to give glyceraldehyde **33**,[146] which, upon isomerisation to dihydroxyacetone **36** and aldol reaction with glycolaldehyde **32**, forms ribose *ribo*-**30**, along with all the other pentose diastereomers.[2] (Some maintain that the pentose products are formed from the addition of the enolate of glycolaldehyde **32** to the carbonyl group of **33**,[147,148] but in fact crossed aldolisation between dihydroxyacetone **34** – due to its reduced hydration – and glycolaldehyde **32** – a good aldol electrophile – is favoured.[2,149]) The formose reaction is essentially an uncontrolled network of aldol condensation, Cannizzaro disproportionation and Lobry van

Troostenburg de Bruyn-Alberda van Ekenstein reactions, however, and all sugars generated are also capable of isomerisation and further reaction. The reaction mixture is further complicated as the sugars are destroyed under the reaction conditions. The exceedingly complex mixture of sugars produced – with no selectivity shown for ribose *ribo*-**30** – are hard to separate, meaning it is difficult to determine a yield of *ribo*-**30** (but thought to be <1%).[150] The slow rate of the initial step also decreases the overall yield of sugar products, as formaldehyde **1** undergoes destructive Cannizzaro disproportionation to methanol and formic acid **2**.[150]

A phosphorylated variant of this reaction has been successful in that it results in a simplified range of reaction products, which, due to their phosphorylation, are more stable under the reaction conditions (***Figure 1.14***).[151] The homoaldol reaction of glycolaldehyde phosphate **32**-P (formed from the reaction of glycolaldehyde **32** with amidotriphosphate **35**[152] or diamidophosphate DAP[153,154]) in aqueous NaOH forms a mixture of tetrose-2,4-diphosphates **39** and hexose-2,4,6-triphosphates **40** in a ratio of 1:10. *Rac*-allose-2,4,6-triphosphate is the major product, in up to 40% yield (the sugar phosphates are formed in up to 80% yield overall). The addition of 0.5 equivalents of formaldehyde **1** lowers the overall yield of phosphorylated products to 45%, but the pentose-2,4-diphosphates **41** (from the crossed aldol reaction of glycolaldehyde phosphate **32**-P and glyceraldehyde-2-phosphate **33**-2P) are produced alongside the hexose-2,4,6-triphosphates **40** (3:1). The formation of *ribo*-**41** is kinetically favoured (making up 52% of all pentose-2,4-diphosphates), but is not the most thermodynamically diastereomer, isomerising to the arabino-derivative over time.[151] Nevertheless, the phosphorylated products are constitutionally stable under the reaction conditions, and the presence of a phosphate group at C-2 of the intermediates and products prevents aldose-ketose tautomerisation *via* the Lobry van Troostenburg de Bruyn-Alberda van Ekenstein transformation, thereby reducing the variety of products.

The pentose-2,4-diphosphates **41** can also be produced from the condensation of glycolaldehyde phosphate **32**-P and glyceraldehyde-2-phosphate **33**-2P in the presence of metal hydroxide minerals.[155] The anionic phosphates are strongly absorbed between the positive layers of the metal hydroxide minerals, thereby concentrating the reactants from dilute neutral solution and inducing aldol condensation. Use of hydroxyapatite (a mineral containing phosphate and calcium

ions) as a catalyst favours the formation of *ribo*-**41** over the other aldopentoses, albeit in only 0.3% yield.[156] There is a major problem with this synthesis, however: pentose-2,4-diphosphates **41**, despite a superficial constitutional similarity to ribose found in the RNA backbone, do not display a natural phosphorylation pattern or level; in particular, the phosphorylation of the C4-hydroxyl prevents the formation of a furanosyl ring.

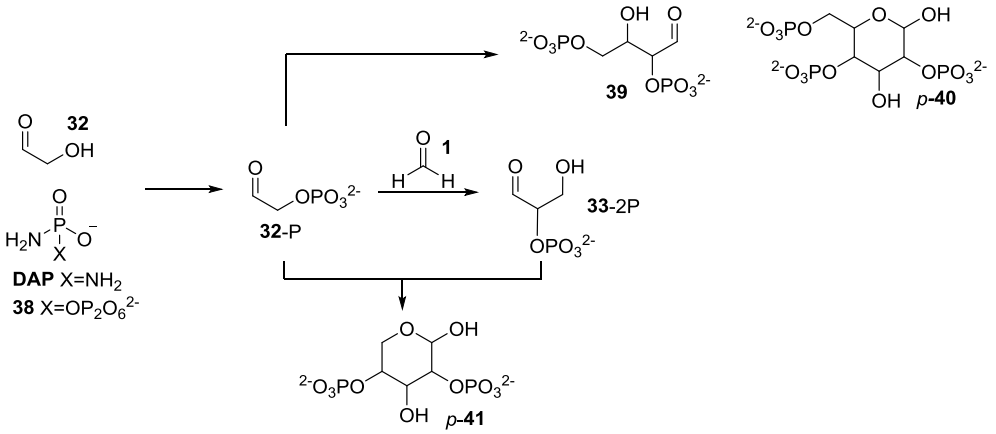


Figure 1.14**: α-Hydroxyaldehyde phosphorylation restricts other chemistries and favours aldolisation. **32** undergoes efficient and selective α-phosphorylation by amidophosphates **38** or DAP to provide glycolaldehyde phosphate **32**-P. Homoaldol condensation of **32**-P provides a mixture of **39** and **40**. In the presence of **1** (a good aldol electrophile), **32**-P preferentially undergoes crossed aldol condensation to afford **33**-2P (another good aldol electrophile, possibly due to an intramolecular H-bond between the OH and CO moieties) which reacts with **32**-P, giving **41.

Once formed, the (in)stability of ribose causes yet further problems. Ribose isomerises in neutral solution at room temperature to a mixture of pentoses, predominantly arabinose, and can then undergo further degradation pathways.[150,157] Ribose has a short half-life, even in neutral solution: 300 days at 25 °C in neutral solution, decreasing to 73 minutes at 100 °C (the other pentoses have slightly longer half-lives, but not significantly different).[158] Phosphorylated sugars are even less stable at pH 7. Ribose-5-phosphate, for example, has a half-life of just 7 minutes at 100 °C, and ribose-2,4-diphosphate has a half-life of 31 minutes at the same temperature (although it is significantly more stable under alkaline conditions).[158] The instability of ribose and other sugars implies that their accumulation for use as prebiotic feedstocks is unlikely.

A suggested solution to this problem is borate complexation. The *cis*-diol of ribose means that it is able to co-ordinate borate. It is known that ribose is stabilised in the presence of borate:[159,160] of all the aldopentoses, ribose has the strongest binding

affinity with borate, and its stability increased the most in the presence of borate,[159] although the increase in stability is modest,[119] and a much greater degree is conferred upon ribulose upon borate-complexation.[148] A natural extension of the exploration of this phenomenon is the investigation of the effect of borate on the formose reaction itself. When heated in alkaline solution (45 °C, pH 12) in the presence of borate minerals, glycolaldehyde **32** and glyceraldehyde **33**, which would otherwise form a complex mixture of sugars and their degradation products, instead from pentose-borate complexes relatively cleanly (***Figure 1.15***).[147] The authors suggest that **33** forms an anionic complex with borate, which, by virtue of its negative charge, would be deactivated to enolate formation but would remain susceptible to nucleophilic attack by glycolaldehyde. The pentose sugar thus formed would be stabilised by *cis*-diol complexation with borate. When using formaldehyde **1** as the initial starting material, **33** would be the first intermediate able to co-ordinate borate (as neither **1** nor **32** contain a *cis*-diol), and so it has been suggested that the first steps of the formose reaction would proceed unhindered in the presence of borate minerals.[147] However, co-ordination of borate to a *cis*-diol-containing intermediate essentially prevents formose cycling, leaving **1** prone to Cannizzaro disproportionation, although this can be mitigated by including an excess of **32** over **1**.[148] Moreover, complexation of **33** with borate prevents isomerisation to dihydroxyacetone **34**, which has been suggested as the key intermediate in the formation of higher-order sugars.[156] The availability of borate on the early Earth in sufficient concentration is also debatable.[161]

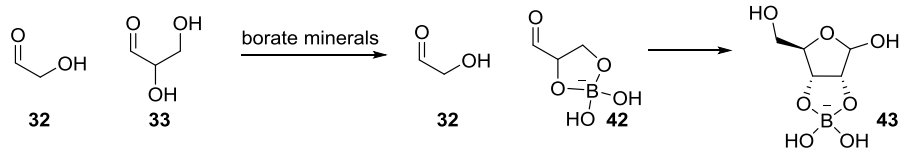


Figure 1.15*: Proposed borate-directed synthesis of borate-stabilised ribose-derivative **43**.In the presence of borate minerals, glyceraldehyde **33** forms anionic complex **42**, but **32** does not co-ordinate borate. It has been proposed that nucleophilic attack of **32** on **42** results in the formation pentose sugars, including **43**; the negative charge of **42** prevents enolate formation and therefore the formation of a variety of by-products.*

1.4.3.2 Prebiotic synthesis of nucleobases

Prebiotically plausible syntheses of the nucleobases have been studied in detail since their discovery in the 1960s.[84] Oró showed that the pentamerisation of hydrogen cyanide in the presence of ammonia at 90 °C results in the formation of adenine **44**, *via* the key intermediate 4-aminoimidazole-5-carbonitrile (AICN, **45**), albeit in very low yield

(***Figure 1.16***).[162–164] Investigations by Orgel and colleagues elucidated some mechanistic details and the chemistry of the intermediates. Polymerisation of HCN forms the relatively stable diaminomaleonitrile (DAMN, **46**),[165] which can be converted to AICN **45** by reaction with (relatively high concentrations of[166]) formamidine **47**[165] or by photochemical rearrangement.[166] The reaction of **45** with further **47**[165] or HCN[167] results in the formation of **44**; hydrolysis of **45** produces 4-aminoimidazole-5-carboxamide (AICA, **48**), which can react with cyanate **49** to give guanine **50**.[167] Problems arise when attempting to account for the accumulation for enough HCN – a volatile compound – for this synthesis: it cannot simply be concentrated in a body of water by evaporation, for example. It has been suggested that nucleobase synthesis could occur in eutectic solution: increased rates of **46** formation due to the effect of HCN concentration during freezing have been observed,[168] and, moreover, this method of accumulation would be compatible with photochemical irradiation required for conversion to **45**.[169] The requirement for formamidine **47** is also an issue – it is unlikely that a sufficient concentration of ammonia for its synthesis were available on the early Earth[13] – but this can be mitigated through the photochemical conversion of **46** to **45**, and the addition of HCN to **45** to yield adenine **44**.

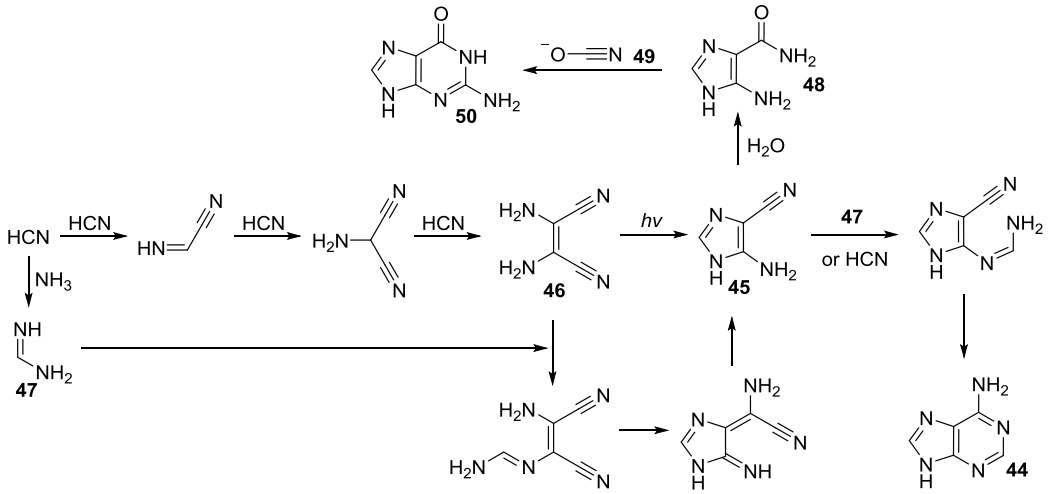


Figure 1.16*: Prebiotically plausible routes to adenine **44** and guanine **50** from hydrogen cyanide. Key intermediate AICN **45** can be accessed from the reaction of DAMN **46** with formamidine **47**, or from the UV irradiation of **46**.*

Orgel and co-workers have also established a synthesis of cytosine **31**, and, by its hydrolysis, uracil **51**, from the reaction of cyanoacetylene **5** and aqueous cyanate **49** (***Figure 1.17***).[170] Cyanovinylurea **52**, a stable intermediate, can be formed in up to 19% yield from the reaction of cyanoacetylene **5** and cyantate **49** at 100 °C and pH 8, which

then rapidly converts to **31** (95% yield) at high temperature and pH, or more slowly at lower temperature or less alkaline conditions. An alternative route involves the reaction of cyanoacetaldehyde **53** (from the hydrolysis of **5**) and urea **54**,[171] thereby avoiding the need to accumulate **49** and **5** in aqueous solution, in which they are both prone to hydrolysis.

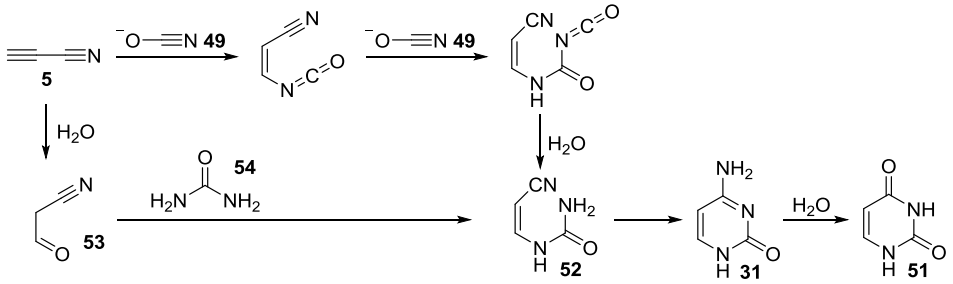


***Figure 1.17**: Proposed prebiotic syntheses of cytosine **31** and uracil **51** from cyanoacetylene **5**.*

1.4.3.3 Glycosylation

Fundamentally different types of chemistries required to build nucleobases and ribose.[119,150] Separate oligomerisations of formaldehyde **1** (to ribose *ribo*-**30**) and cyanide (to adenine **44**) in the same environment are prevented by glycolonitrile formation. It has been suggested that hydrogen cyanide could be sequestered as ferrocyanide,[2] but this still necessitates the spatial and temporal separation of the two syntheses. *Ribo*-**30** and the nucleobase must then be brought into the same environment for glycosylation to occur. The difficulty of forming a glycosidic bond between the sugar and the nucleobase means that the synthesis of nucleotides from these components in significant yield is near-impossible.

The thermodynamic and kinetic barriers to this reaction are summarised in a review by Sutherland:[91] *ribo*-**30** exists in an equilibrating mixture of furanosyl and pyranosyl isomers and their various anomers, and so only a small amount of the desired α-furanose form is present. The nucleobases are also present as a mixture of tautomers, with various nitrogen atoms the most nucleophilic. Activating *ribo*-**30** under acidic conditions protonates the nucleobase, removing its nucleophilicity. Moreover, hydrolysis of the glycosidic bond is favoured over its formation.

Assuming the components of the conceptually simplest retrosynthesis of ribonucleotides can be formed prebiotically, the stereospecific coupling of the bases to the anomeric position of the sugar remains to be accomplished. Having failed to

achieve glycosylation in aqueous solution, Orgel demonstrated that ß-adenosine ß-**55** (in a yield of just 4%), ß-guanosine ß-**56** (9%) and ß-inosine ß-**57** (8%) could be obtained by heating ribose and the respective nucleobase in the dry state in the presence of magnesium chloride and sodium trimetaphosphate **58** (***Figure 1.18***).[172] A variety of isomeric by-products were also formed (disappointingly, although ribose *ribo*-**30** was partially phosphorylated, no phosphorylated adenosine derivatives were produced).

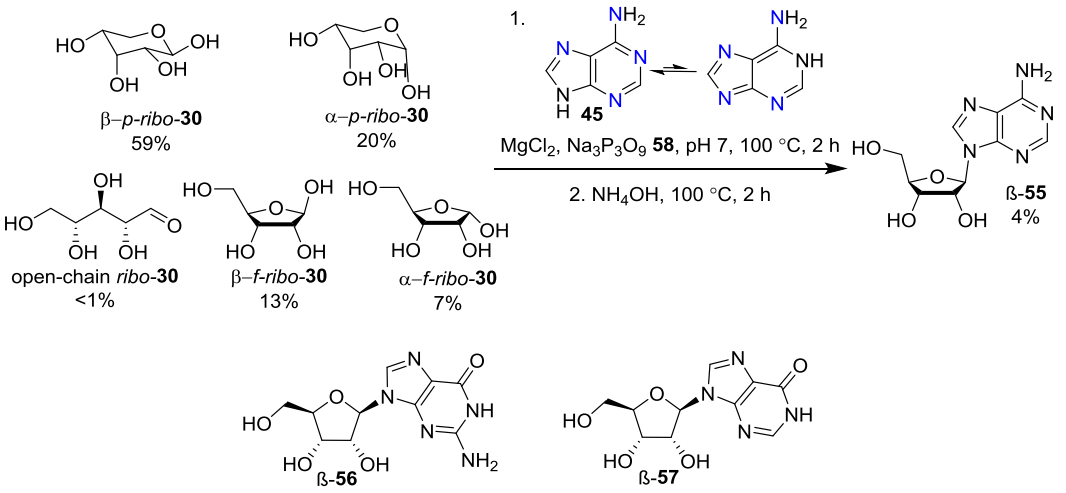


Figure 1.18*: Synthesis of adenosine ß-**55** by glycosylation of ribose ribo-**30** (showing the anomers present in equilibrium in aqueous solution) and adenine **44** (the tautomers present in aqueous solution are shown, with the potential nucleophilic sites highlighted in blue). The structures of ß-guanosine ß-**56** (9%) and ß-inosine ß-**57** are also shown.*

The prospect for pyrimidine synthesis is even worse: the *N*-1 lone pairs of cytosine **31** and uracil **51** are delocalised around the aromatic ring and into the carbonyl bond, with the same thermodynamic barrier remaining. Therefore, the synthesis of a glycosidic bond formed experimentally between *N*-1 of **31** and *ribo*-**30** has not been reported.

Activation of ribose by other means has been suggested to promote glycosylation reaction. It has been proposed that complexation of ribose with borate facilitates purination, but experimental evidence to support this theory is lacking.[173] The proposed mechanism[173] is similar to the extant biosynthesis of nucleosides which uses a pyrophosphate leaving group at the anomeric position.[174] Use of a phosphate leaving group has been partially successful in promoting glycosylation under prebiotically plausible conditions: although the reaction of ribose-1-phosphate with a nucleobase was unsuccessful,[13] reaction of 5-phosphoribosyl-1-pyrophosphate with hypoxanthine

has reportedly generated inosine-5′-monophophate in yields of up to 20%.[174] Reaction of ribose-1,2-cyclic phosphate with adenine or hypoxanthine results in the formation of the ß-nucleotide-2′-phosphate in respectable yields (8-15%).[175]

1.4.3.4 Phosphate

The importance of phosphate in life today is undisputed. Phosphorus, as phosphate, is entirely ubiquitous throughout life: nucleic acid backbones are made up of phosphodiester linkages; cell membranes contain phospholipids (molecules with long hydrophobic hydrocarbon tails and hydrophilic phosphate-containing heads); and many phosphorylated biomolecules and intermediates are found in metabolism, most notably in ATP, which stores and releases chemical energy by virtue of its phosphate linkages. The topic of incorporation of phosphate during the abiogenesis of biomolecules is therefore an important one, but the prebiotic availability and chemistry of phosphate has been much disputed.

A seminal article by Westheimer, published 30 years ago, rationalised the advantages the presence of phosphate-containing biomolecules to life.[176] The importance of ionised metabolites had already been noted: ionisation of a molecule decreases its permeability across a lipid membrane, allowing it to be retained within a cell.[177] The first pK_as of phosphoric acid and its derivatives (phosphate mono- and diesters) are around 2, so phosphates are ionised at physiological pH and trapped within the cell.[176] The valency of phosphate is also important, allowing it to form two ester linkages (thereby acting as a nucleic acid backbone) whilst retaining a negative charge on a third oxygen. This negative charge solubilises the biomolecule in aqueous solution, but also repels approaching nucleophiles, meaning that phosphate esters are inherently stable compared to other esters.[176] The negative charge imposes a high energy barrier for hydrolysis; a large amount of energy is released as a result of organophosphate cleavage which can be harnessed to drive metabolic reactions.[178] The relatively high kinetic stability and thermodynamic instability of organophosphates can essentially be fine-tuned depending on the exact local environment (such as by charge manipulation in an enzyme active site).

1.4.3.4.1 Sources of phosphorous

The ubiquity and centrality of phosphate in extant life makes it seem obvious that chemistry occurring at the origins of life incorporated phosphorous. The biological

ubiquity of phosphorus suggests it would have been readily available on the early Earth,[144] but a prebiotic source of phosphate (especially in a suitably reactive form) is debated.[179] Although phosphorus is relatively common element in the Earth's crust, it is mostly in the form of the minerals apatite and fluoroapatite.[180] The low solubility of these phosphate minerals, and the tendency of phosphate to precipitate from solution in the presence of abundant divalent metals such as Ca^{2+} and Fe^{2+}, means a likely low concentration of dissolved phosphate in the prebiotic ocean.[180] Extant organisms are capable of collecting, storing and using phosphorus from such dilute solutions,[179] but it is inconceivable that the earliest forms of life were also capable of this.[181] This has led to the suggestion that biology evolved to use the ideal phosphate-containing system, but that the earliest forms of life were based on a different, non-phosphorylated genetic system.[182] The synthesis and investigations into the properties of such genetic systems are discussed in further detail later in this chapter.

Abiotic phosphate sources other than insoluble terrestrial apatites have been suggested.[180,183,184] In particular, phosphides such as schreibersite ($(Fe,Ni)_3P$), which may have arrived on Earth on iron meteorites during the Late Heavy Bombardment, could act as a source of reactive reduced phosphate.[183,184] Schreibersite has been shown to react directly with glycerol to form phosphite and glycerol phosphate[185] (a component of phospholipid membranes), but phosphite is also formed upon the reaction of schreibersite with water.[183] Phosphite is both a thousand times more soluble in water, and more reactive, than phosphate, and therefore this could act as a supply of dissolved phosphorus in the early oceans.[183] The presence of phosphite in early Archean marine carbonates indeed suggests that phosphite would have been an abundant species in oceans before 3.5 Gya.[185] In addition to phosphite, the corrosion of schreibersite in water also leads to the formation of orthophosphate, hypophosphate and pyrophosphate, probably *via* radical recombination mechanisms.[186]

Proposals for the mitigation and circumvention of phosphate precipitation have also been made. The corrosion of schreibersite also leads to the formation of insoluble vivianite ($Fe_3(PO_4)_2 \cdot 8H_2O$), and it has been proposed that a high cyanide concentration could trap Fe^{2+} ions of vivianite as hexacyanoferrate (II), thereby freeing phosphate.[64] Hydrogen sulfide and sulfur dioxide are capable of solubilising phosphates, so it is easy to envision local environments in which apatite is more easily solubilised leading to a high phosphorus concentration.[179] The continuous production and cycling of phosphate

could have been possible on a more volcanically active early Earth: P_4O_{10} can be volatilised from rocks under volcanic conditions, forming orthophosphate and a mixture of polyphosphates upon rapid cooling and partial hydrolysis after eruption.[187]

1.4.3.4.2 Phosphorylation

Orthophosphate is inherently stable, and does not readily undergo condensation reactions to form phosphoesters under ambient conditions. A further problem comes from the perceived ubiquity of water as a prebiotic reaction solvent: condensation reactions are generally thought to be thermodynamically disfavoured in water.[144] One approach has been the use of suitable activating agents:[13] generally, an activated phosphate is generated by the attack of the nucleophilic phosphate group on an electrophilic species. Cyanovinylphosphate **59** is one such example, formed from the reaction of phosphate and cyanoacetylene **5,** the latter of which has been implicated in other areas of prebiotic chemistry.[188] Cyanovinylphosphate **59** is relatively stable in water, and has been shown to produce uridine-5′-monophosphate **60**-5′P in 4% yield (based on **59**, used in two-fold excess) from uridine **60** after heating at 60 °C for 18 days in neutral aqueous solution (***Figure 1.19***).[188]

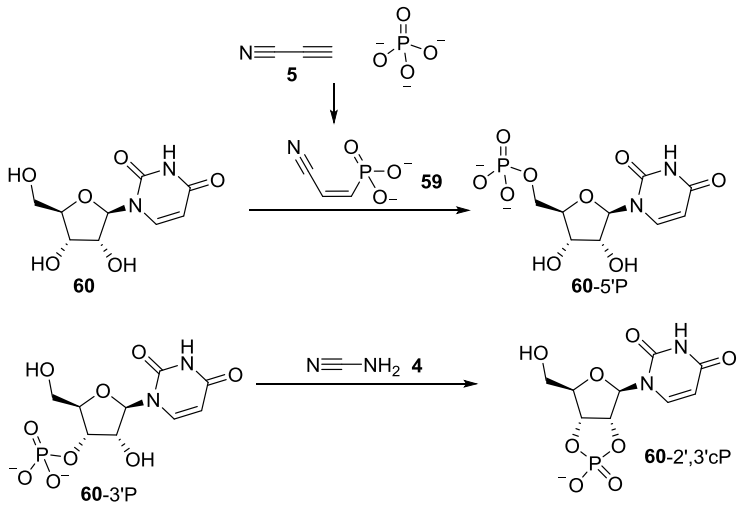


Figure 1.19*: Aqueous phosphorylation of uridine **60** and phosphorylated derivatives. **Top**: **59** produces **60**-5'P from **60**, albeit in low yield. **Bottom**: **4** can be used as an activating agent, promoting the cyclisation of **60**-3'P to provide **60**-2',3'cP.*

Various other prebiotically plausible condensing agents, such as cyanamide **4,** cyanate **49** and cyanoformamide, have also been shown to effect this transformation.[189] The activation and intramolecular esterification of a tethered phosphate group proceeds rather more efficiently: for example, cyanamide **4** has been shown to effect the

formation of uridine-2′,3′-cyclic monophosphate **60**-2′,3′cP from uridine 2′- and 3′-monophosphates in 45% yield when heated at 37 °C, pH 6; or in 73% yield when heated at 65 °C, pH 5.[189] The high effective molarity of the free *cis*-hydroxyl group outcompetes attack of water on the activated phosphate: the low yields of phosphorylation (opposed to cyclisation) achieved by activated phosphates are likely due to competing hydrolysis, producing orthophosphate, rather than inefficiency of the reaction mechanism.

To mitigate for the solubility problem, has been suggested that reactions involving the use of activating agents could occur heterogeneously on the surface of phosphate-containing minerals,[189] but this would restrict the formation of nucleotides by this method to suitable locations. The phosphorylation of free ribose reportedly occurs more efficiently than that of the nucleoside under the same conditions (due to the reactive hemiacetal moiety), generating ß-ribofuranose-1-phosphate in up to 20% yield, with no other phosphorylated isomers reportedly detected.[190] The spontaneous formation of ribose-1-phosphate (6% at room temperature) from ribose and phosphoric acid trapped in microdroplets has also been reported: microdroplets (which can be found in clouds and formed by breaking waves) often feature chemical reaction properties not observed in bulk solution.[191]

One way in which the thermodynamic barrier imposed by the presence of water, and the competing reaction with water, in condensation reactions with phosphate could be circumvented is by carrying out reaction in the dry state. Prebiotic environments in which this could occur are easy to envisage, and dry-wet cycles have been proposed for a number of prebiotic processes,[77] providing both the extreme dehydrating conditions to facilitate condensation, followed by increased mobility in a fluidic phase. Heating a nucleoside with inorganic phosphate in a thin film has been shown to result in the formation of nucleotides.[192,193] Uridine monophosphates can be formed in 16% yield from uridine **60** and $NaH_2PO_4 \cdot H_2O$ after 2 h at 160 ºC; lower temperatures (85 °C) result in much slower reactions.[192] More acidic phosphates, such as $Ca(H_2PO_4)_2$, have been noted to provide higher yields than less acidic counterparts (such as NaH_2PO_4).[193] Soluble acidic forms of phosphate, however, are uncommon, and although it has been suggested that a suitably acidic environment for phosphorylation could be formed when ammonium phosphate is heated (upon which ammonia is

lost),[193] it is not clear how ammonium phosphate would form, as the magnesium and calcium phosphates would precipitate before ammonium phosphate.[194]

The addition of urea **54** and ammonium chloride allows phosphorylation in high yield using neutral or basic forms of phosphate: nucleotides can be generated in up to 96% yield (as a mixture of isomers) when heated as a thin film at 100 °C for one day.[195] Omission of ammonium chloride leads to less extensive phosphorylation; phosphorylation is negligible in the absence of urea **54**. The exact role of **54** in this process is not known: it is possible that it acts as a pseudo-solvent at the elevated temperature of the reaction, but it has also been suggested that it could act as a nucleophilic activating agent.[138] Nucleophilic attack of urea **54** on acidified phosphate could displace water to yield an electrophilic intermediate, which is subsequently attacked by a nucleoside hydroxyl, regenerating **54** (***Figure 1.20***).

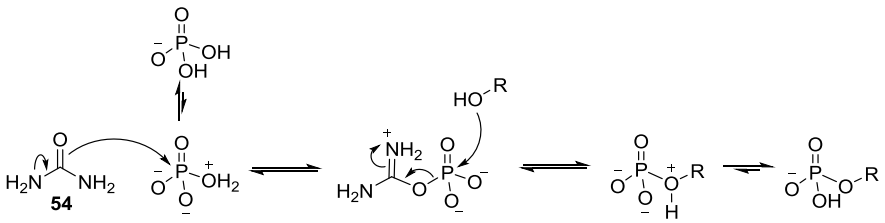


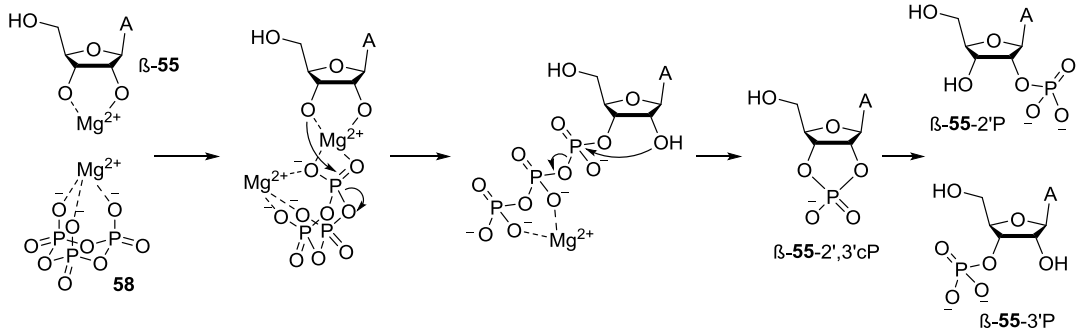
Figure 1.20*: Possible mechanistic role of urea* ***54*** *as a nucleophilic activating agent in the dry state phosphorylation of alcohols.*

It has also been shown that a eutectic solution of urea/ammonium formate/water (1:2:4, although the composition of the mixture changes upon heating, partially converting to formamide **3**) in simulated wet-dry cycles is able to mobilise phosphate from hydroxyapatite in the presence of $MgSO_4$.[196] When heated at 85 °C under these conditions, adenosine monophosphate is produced in around 4% yield from adenosine and hydroxyapatite; 21% yield from Na_2HPO_4; and 15% from struvite ($MgNH_4PO_4 \cdot 6H_2O$). Struvite can also phosphorylate nucleotides (to give nucleotide diphosphates, containing a pyrophosphate group) when heated in anhydrous conditions with urea **54**, although free pyrophosphate is also produced.[194] Struvite has been proposed as a more prebiotically plausible form of ammonium phosphate[194] that precipitates from soluble phosphates in the presence of >0.1 M ammonia,[197] but this also requires a high phosphate concentration, and therefore its likely presence on the prebiotic Earth is disputed.[198] Luneburgite (a mineral containing magnesium, borate and phosphate) has also been proposed as a prebiotically plausible phosphate, purportedly formed from the sequestration of phosphate by borate and Mg^{2+} in the

presence of calcium.[199] Uridine-5′-phosphate (15%) is formed when uridine is heated with in urea **54**; in water, uridine extracts borate, by *cis*-diol complexation, from the mineral with concomitant phosphate release.[199]

The low reactivity of orthophosphate in aqueous solution could also be circumvented by using more energetic condensed phosphates. The low solubility of terrestrial sources of phosphate, however, and the inactivity of orthophosphate, the presumed precursor of condensed phosphates, means that the prebiotic availability of polyphosphates for phosphorylation reactions has also been questioned. The repulsion between two anionic phosphate groups across a phosphoanhydride linkage drives bond cleavage, and the lower pK_a values of phosphate and polyphosphates compared to hydroxide, makes them better leaving groups.[178] Polyphosphates have been detected among the products of oxidation reactions of phosphites[200] and in the reaction of phosphates with urea **54**.[194,197] The linear polyphosphates have not been shown to be effective phosphorylating agents in aqueous solution, but they are easily converted to cyclic trimetaphosphates in the presence of divalent metal ions.[13]

The chemistry of cyclic trimetaphosphate **58** is of particular note, and it has been extensively investigated as a prebiotic phosphorylating agent. It has been observed as one of the products of volcanic P_4O_{10} hydrolysis,[187] phosphite oxidation,[200] and, along with other polyphosphates, from heating ammonium dihydrogen phosphate in urea **54**.[201] It is highly soluble in water and reacts with nucleosides to form a mixture of 2′- and 3′-phosphates in yields of up to 90% (depending on the identity of the nucleoside and of the phosphate's counter-ion) in alkaline solution.[202,203] Much lower yields are achieved at lower pH, but 2′,3′-cyclic nucleotides are also formed:[203] it is likely that the cyclic phosphate is the product of this phosphorylation, but is hydrolysed at high pH (***Scheme 1.21***). The use of a Mg^{2+} catalyst can increase the yield of phosphorylated products seen in neutral solution: cyclic-2′,3′-adenosine monophosphate (3.8%) as well as small amounts of ATP (<0.1%) have been observed.[204] **58** is incapable of phosphorylating deoxyribonucleotides under the same conditions, demonstrating the necessity of a 2′-hydroxyl group to the reaction mechanism.[205] Ca^{2+} is not capable of catalysing this phosphorylation, and it has been proposed that Mg^{2+} forms a chelating complex with the 2′,3′-diol of the nucleoside and **58**, but the larger radius of Ca^{2+} prevents this.[204]



***Scheme 1.21**: Possible mechanistic role for catalytic Mg^{2+} in the phosphorylation of ß-**55** with **58**, forming ß-**55**-2',3'cP which subsequently hydrolyses to ß-**55**-2'P and ß-**55**-3'P. For clarity, only nucleophilic attack of 3'-OH on **58** is shown, but 2'-OH can also behave as a nucleophile.*

Cyclic trimetaphosphate has proved an ineffective phosphorylating agent for sugar *cis*-diols such as glycolaldehyde, however, as the alkaline conditions required result in homoaldol degradation of the starting material.[152] Amidophosphates, in contrast, have been shown to efficiently and regioselectively phosphorylate α-hydroxyaldehydes.[152–154] Ring-opening addition of ammonia to cyclic trimetaphosphate **58** results in the formation of the linear polyphosphate amidotrimetaphosphate **38**,[206,207] which readily adds to the carbonyl group of α-hydroxyaldehydes at neutral pH to form hemi-aminals (***Figure 1.22***). When a stoichiometric amount of magnesium is present, the hemi-aminal undergoes ring closure with loss of pyrophosphate to provide a cyclic phosphoramide **61**, and subsequent hydrolysis and loss of ammonia from the phosphate forms the aldehyde-2-phosphate in quantitative yield.[152] Tethering the phosphate results in a high effective molarity of the hydroxyl group, meaning that the rate of the otherwise slow nucleophilic attack of the hydroxyl on the phosphate group is increased. The magnesium ion is thought to bind pyrophosphate, making it a better leaving group and removing the requirement for alkaline-promoted hydroxyl nucleophilicity, although it is also possible that Mg^{2+} chelation aids the correct orientation of the reaction centres. Diamidophosphate DAP, formed from further ammonolysis of amidotriphosphate **38**,[207] displays similar reactivity towards α-hydroxyaldehydes but without the need for magnesium. Glycolaldehyde phosphate **32**-P, formed from the phosphorylation of glycolaldehyde **32** with either of these reagents, has been suggested as a building block for (relatively) stable diphosphophosphorylated pentose sugars,[151,155] although these display unnatural phosphorylation patterns. Nevertheless, if phosphorylation prevents these sugars from being used in nucleotide synthesis, they are still suitable for incorporation in other metabolic pathways.[154]

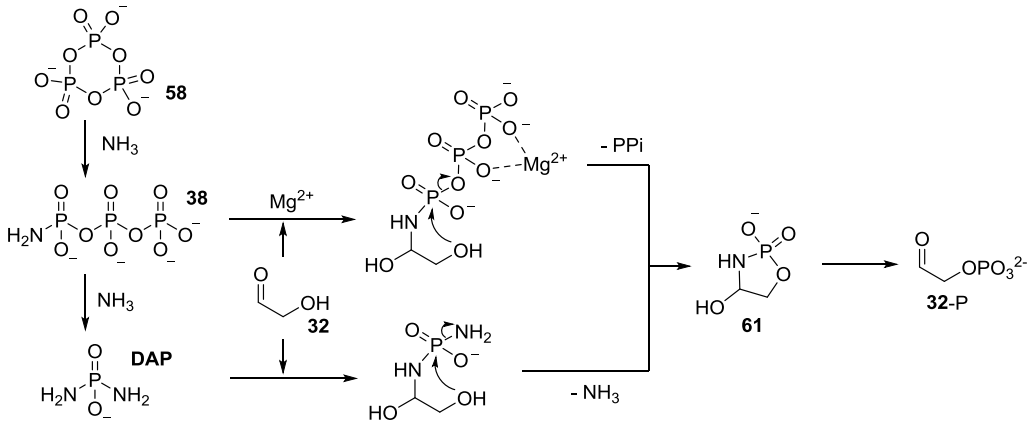


***Figure 1.22**: The formation of amidotrimetaphosphate **38** and DAP from the ammonolysis of cyclic trimetaphosphate **58**, and their use in the phosphorylation of glycolaldehyde **32**.*

Both amidotrimetaphosphate **38** and DAP can be used for the phosphorylation of tetrose and pentose sugars, forming 1,2- and 2,3-cyclic phosphates which hydrolyse upon exposure to acidic conditions to yield sugar-2- or -3-phosphates.[153] DAP has also been used for the phosphorylation of nucleosides and nucleotides: forming the nucleotide-2′,3′-cyclic phosphate when nucleosides or 2′-or 3′-nucleotide monophosphates are used, and the corresponding 5′-nucleotide-amidodiphosphates from nucleotide-5′-monophosphates.[208]

1.4.4 Why RNA?

From this evidence, it appears that the assembly of RNA was not easy to accomplish on the prebiotic Earth, with yields diminishing at each step – the synthesis of ribose leads to a complex mixture of sugars which is incompatible with nucleobase abiogenesis, syntheses of the nucleobases are low yielding and incompatible with sugar synthesis, phosphorylation leads to a mixture of structural isomers, glycosylation is inefficient at best and oligomerisation generates a heterogenous backbone – and so RNA would not have been chosen as the system on which life was initially based simply because its formation was facile. Biological function ultimately derives from the properties and therefore chemical structure of nucleic acids. RNA (and DNA after it) must have been chosen because it is biologically superior to the alternatives. It is unclear, however, what constitutes biological superiority in this context.

It is possible that ribose was selected as an important structural component in nucleic acids because it is able to permeate fatty acid and phospholipid membranes faster than other aldopentoses and hexoses (presumably its conformation and flexibility give it a

kinetic advantage when crossing the membrane): before the evolution of protein transporters, this may have conferred a significant selective advantage on protocells containing genetic material based on ribose.[209] It is still important, however, to investigate whether there are any other disadvantages associated with non-ribose based nucleic acids.

The synthesis of unnatural genetic systems allows us to elucidate reasons why nature selected the canonical nucleic acids. Eschenmoser sought to investigate the chemical and physical properties of alternative nucleic acids: in particular, by systematically exploring the properties of nucleic acids in which the 3′,5′-ribofuranose-phosphate backbone was replaced with other structures. By comparing these with the properties of DNA and RNA he could rationalise the relationship between structure and function in nucleic acids. In effect, Eschenmoser was simulating evolution in these experimental studies: the variation of structure and comparison of function could indicate the parameters for selection that may have led to the formation of RNA and its role as genotype and phenotype.

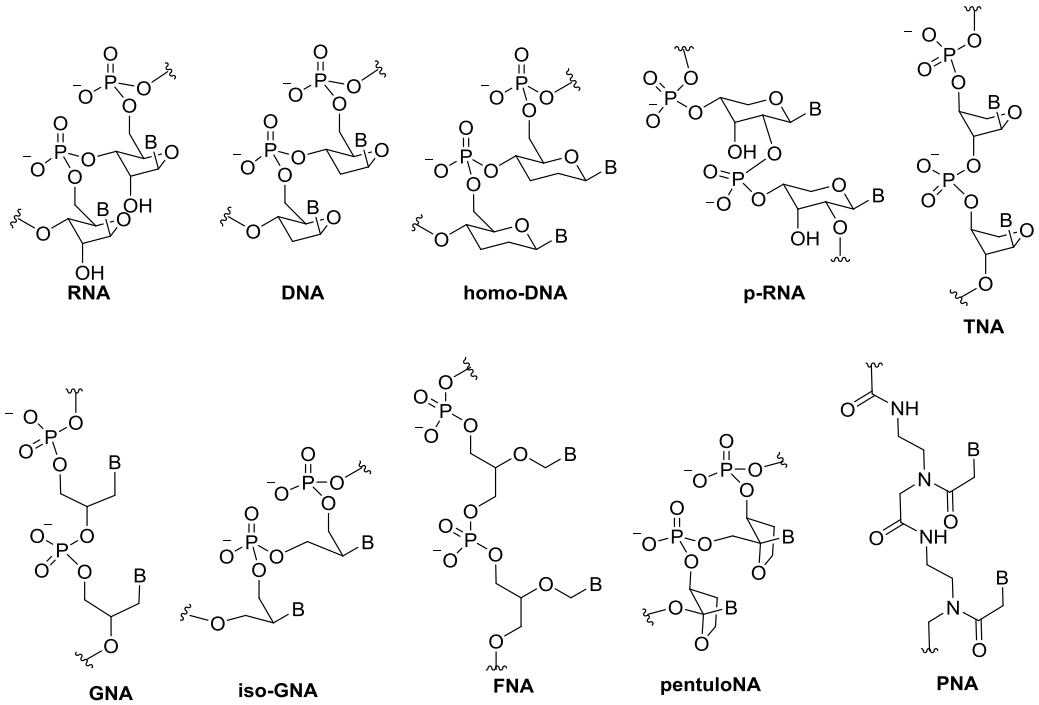


***Figure 1.23**: A selection of XNAs used to reveal the relationship between structure and function in nucleic acids.*

Investigations into the aldomerisation of glycolaldehyde phosphate under basic conditions, forming phosphorylated hexose (in the absence of formaldehyde) and

pentose (in the presence of formaldehyde) sugars,[151] led Eschenmoser to ask why Nature had selected pentose and not hexose sugars to build her genetic materials, when the latter, from a chemical point of view, appeared to be an equally valid candidate.[94,210,211]

Eschenmoser therefore sought to investigate the properties of an unnatural nucleic acid with a six-membered ring backbone, using, for the ease of synthesis, a model nucleic acid system based on 2′,3′-dideoxyglucopyranose (named homo-DNA) for initial investigations.[212] This is constitutionally more similar to natural DNA (it has only a single extra CH_2 group within the ring, compared to an extra CH_2 and two phosphate groups in an allose-2,4,6-triphoshate system), meaning that the properties conferred by a change in ring size could be directly compared. The synthesis of homo-DNA[213] revealed for the first time that an unnatural system was capable of Watson-Crick base-pairing with a complementary oligomer of homo-DNA,[212,214] but formed a more stable duplex than the same sequence with a DNA backbone,[214] thought to be due to the rigidity of the pyranosyl-based backbone compared to one that is furanosyl-based.[210] This links the relative stabilities of base-pairs to the structure of the sugar backbone, demonstrating that base-pairing is not reliant on the nature of the nucleobase alone. Comparing the base-pairing stability of homo-DNA (a dideoxy-pyranose ring) with allo-, altro- and glyco-pyranosyl 6′,4′-oligonucleotides (all fully hydroxylated pyranoses) indicates that steric hindrance in the fully hydroxylated hexopyranosyl oligonucletides inhibits base pairing.[215] This was confirmed by the elucidation of crystal structure of the homo-DNA duplex, which indicated that fully hydroxylated hexose sugar rings sterically hinder the base from orientating in a way that allows Watson-Crick base pairing and base staking.[216–218] It was therefore proven that a functional nucleic acid must be based on pentose not hexose, although it was not clear why extant nucleic acids are based on furanose not pyranose sugars.

Synthesis of the pyranosyl form of RNA (containing 2′,4′-linkages, p-RNA) revealed that it shows stronger base-pairing in duplexes than either DNA or RNA duplexes, and that base-pairing is confined to the Watson-Crick mode.[219,220] Maximal base-pairing strength, however, is not necessarily optimal base-pairing strength: strong base-pairing may be beneficial for template-directed ligation, but would hamper strand separation required for replication. In marked contrast, the 3′,4′-isomer of p-RNA shows no base-pairing ability – possibly because the axial-equatorial phosphodiester bridge is

shortened from six to five covalent bonds – although the lyxo 3′,4′ (diaxial) isomer does display base-pairing.[221] This indicates that nucleic acids based on shorter backbones, such as tetroses, may be functional.

Erythrose-2,4-diphosphate and threose-2,4-diphosphate are formed in the aldol condensation of glycolaldehyde phosphate,[151] which suggests that the prebiotic formation of threose nucleic acids (TNA) is possible. Synthesis of TNA[222] allowed its properties to be experimentally explored: TNA displays efficient base-pairing similar to RNA with respect to specificity, strand orientation and pairing strength,[223] and is also capable of cross-pairing with RNA and DNA[224] (the properties of TNA with a phosphoramidate linkages is not dissimilar[225]). The lack of a 2′-OH, however, reduces the phenotypic capacity of TNA compared to RNA.[226]

The acyclic equivalent of TNA, glycol nucleic acid (with a 3′,2′-linkage, GNA), has a high duplex stability – surprising, given its constitutionally simple backbone.[227,228] GNA nucleotides can be formed from the regioselective and stereospecific nucleophilic ring-opening of an epoxide,[227] or from the acid-catalysed substitution of one of the primary hydroxyls of glycerol with a nucleobase[229] (glycerol has been found on meteorites[230] and therefore may have been available on the early Earth for incorporation into proto-biomolecules). A mixture of two complementary GNA single strands can form an anti-parallel duplex of GNA; parallel duplexes do not form, nor do single GNA strands self-pair.[227] Iso-GNA (containing 3′,1′-linkages), however, displays only limited base pairing,[231,232] and another isomer (called glycerol nucleic acid[233,234] or flexible nucleic acid, FNA[231]) shows no base pairing.[233] A shift in the position of the nucleobase on the acyclic backbone (from 1′ to 2′), or a variation in the backbone-nucleobase linker length (the three-bond link in FNA is perhaps too flexible; the one-bond link in iso-GNA results in a shorter interstrand distance and increases electrostatic repulsion; the 2-bond linker in GNA is the Goldilocks distance) changes the base-pairing abilities of the XNA.

The keto sugars appear to be more abundant than the corresponding aldo sugars in many prebiotic scenarios, but ribulo- and xyulo-oligonucleotides show no base-pairing ability: the furanose ring orientates the nucleobase in an *anti*-conformation, with the Watson-Crick hydrogen-bonding face directed towards the sugar-phosphate backbone. The phosphates are therefore placed closer together than in an equivalent RNA duplex; the increased electrostatic repulsion appears to be responsible for the absence

of base-pairing in pentuloNA.[232] The ribofuranosyl ring in RNA also orientates the nucleobase in an *anti*-conformation, but in a way that the Watson-Crick hydrogen-bonding face directed away from the sugar-phosphate backbone.[235] RNA can therefore form a duplex with phosphates on the outside (as far away from each other as possible, minimising electrostatic repulsion) and the nucleobases buried in the hydrophobic core.

The acyclic equivalent of pentuloNA is isoGNA, which, conversely, is capable of (limited) base-pairing:[231] the nucleobases are directly connected to the backbone and can re-orientate so their hydrogen-bonding faces point away from the sugar backbone, allowing greater interstrand distances.[232] The opposite phenomenon (loss of base-pairing) is observed when the ring framework of RNA is removed to give FNA. The nucleobase is connected to the FNA backbone through a flexible linker, which allows the nucleobase to adopt an unfavourable anti conformation, whilst free rotation in the linker allows irregular arrangements of the nucleobase on the backbone.[232]

These investigations demonstrate the importance of the furanosyl ring in RNA: it restricts nucleobase rotation and orientates it in the *anti*-conformation. The Watson-Crick hydrogen-bonding face therefore points away from the sugar-phosphate backbone, ensuring an optimal interstrand distance in a duplex: a shorter interstrand distance results in a weaker-than-RNA duplex, with limited base-pairing ability; a longer interstrand distance results in a stronger duplex that is less sensitive to nucleobase mismatches.[232] Therefore, although it might be assumed that the nucleobase is the most important sub-structure within the nucleic acid, this appears not to be the case.

It is not clear why canonical nucleic acids feature ribose and not the other pentoses. It has been suggested that selection could take place during oligomerisation: the more nucleophilic *cis* 2′,3′-OH groups of ribonucleotides would oligomerise more efficiently than *arabino*-, *xylo*- and *lyxo*- analogues (*lyxo*- also has *cis* 2′,3′-OH groups, but these are more sterically hindered than *ribo*-).[236]

Nucleic acid analogues with phosphate-free backbones have also been synthesised. Nucleic acids with a peptide backbone (peptide nucleic acid, PNA) can self-hybridise to form a PNA duplex,[237] and can form a PNA–DNA hybrid duplex by Watson-Crick base-pairing with a complementary strand of DNA,[238,239] or by the template-directed

oligomerisation of PNA on a strand of DNA.[240] This indicates that the sugar-phosphate backbone is not necessary for the formation of helical duplexes, but PNA exhibits much stronger base-pairing than DNA.[237] Similarly, replacing anionic phosphates in a DNA backbone with uncharged sulfone linkers forms a stronger base-pairing between dimers, but longer oligosulfones fold in on themselves, preventing pairing with a second strand.[233] (The replacement of the C5′-O and C3′-O with methylenes also changes the sugar pucker of the backbone and reduces the solvation of the duplex.[241]) Furthermore, oligosulfones exhibited different physical properties and levels of chemical reactivity when only one nucleobase was changed: the polyanionic backbone of extant nucleic acids dominates their physical properties,[233] meaning that the information it encodes can change, allowing Darwinian evolution, without changing its replicability.

The importance of ionisation of biomolecules,[177] and specifically the role of phosphates,[176] is well-known. It is unusual for one polyanion to bind another, but crucially the repeated negative charge moderates and controls the formation of a duplex. The anionic repulsion within a single strand prevents it from folding, meaning the oligomer adopts an extended conformation that allows templating (although association with cations reduces the electrostatic charge repulsion sufficiently to allow folding and function); repulsion between strands essentially destabilises the helix, so that the duplex formed can disassociate and each strand act as a template for replication. This repulsion also directs inter-strand interactions away from the backbone, so that hydrogen bonding between strands occurs at the Watson-Crick faces of the nucleobases (which are held within the hydrophobic core of the duplex, excluding water which could form competing hydrogen bonding interactions) and accurate information transfer can occur. On the other hand, the repeating anionic phosphate means that inexact copies can also be generated with essentially the same physical properties as the original strand. Together, this allows Darwinian evolution to occur.

Synthesis of nucleic acids containing unnatural nucleobases has allowed the properties of the nucleobase that contribute to base-paring ability to be investigated. Base pairing follows two rules of complementarity: size complementarity (a single ringed pyrimidine with a double-ringed purine) and hydrogen-bonding complementarity (H-bond donors pair with H-bond acceptors), and these two factors contribute roughly equally to base-

pairing strength.[242] Observations that modifications to nucleobases affect the strength of the glyosidic bond when they are incorporated in a nucleoside have led to the suggestion that the canonical nucleobases were selected by evolution for optimal stability of the nucleoside.[243]

Exploration of the base-pairing ability of artificial nucleobases indicated another critical feature of the nucleobases: their relative hydrophilicity and hydrophobicity.[244] In contrast to the nucleic acid backbone, and many other biomolecules, it is important for the nucleobase not to be ionised.[242,244] If the difference between the pH of the environment and the pK_a of the nucleobase is greater than 2, the base remains in its unionised, protonated form. To minimise interactions with the aqueous solvent, the bases instead assemble, an association reinforced by stacking interactions and specific hydrogen bonding and adopting the most stable arrangement: a duplex structure. The canonical nucleobases have pK_as around 4 or 10 (limited evidence suggests that bases with pK_as much lower or higher than these respectively have compromised base-pairing ability[244]). It therefore appears that the natural nucleobases are those that form the most favourable base-pairing interactions, based in the physiochemical properties in water at near-neutral pH. The ability of adenine and cytosine to protonate, thereby shifting their pK_a to near 7 at near neutral pH may have important structural and functional implications: their acid/base catalytic function is increased, and they are able to form a number of different base-pairing modes, increasing the structural diversity of an RNA molecule.

This advantage conferred by the canonical nucleobases appears to be selected for at the level of the oligomer (single nucleobases and nucleotides cannot base-pair). Indeed, each component of RNA appears to be selected through the appearance of functional properties at oligomer level in a duplex: RNA therefore appears to be an emergent structure with emergent properties (the whole is greater than the sum of the parts[245]).[232,235] Furthermore, these properties are optimised in an aqueous environment at neutral pH and near-room temperature. Of course, if the conditions on the early Earth were different, then another system could have been optimal: if the temperature of the environment had been higher, for example, RNA's base-pairing would be too weak, whilst that of p-RNA would have been 'just right'.[235] This observation could be interpreted in two ways: either we can assume that the conditions on Earth today were similar to conditions on the early Earth, and that physical extremes should be avoided

when testing theories of life's establishment in the laboratory, or that the early environment was different to today's, and another genetic system preceded RNA.

The notion that a different nucleic acid analogue would have preceded RNA at the first stages of the evolution of genetic molecules is often referred to as the pre-RNA World theory, and has frequently been raised as a solution by some researchers to the perceived difficultly of synthesising RNA.[84,234,243] Therefore, screening the structural neighbourhoods of the canonical nucleic acids, in addition to improving our understanding of how structure and function are related, has also been carried out with the intention of discovering an alternative genetic system from which RNA could have evolved.[234,246]

It is not clear how the transition between a pre-RNA to an RNA World would have taken place, but the pre-RNA should be capable of storing genetic information and be chemically simpler, or at least more easily formed on the early Earth. A continuous transition,[84] in which genetic information is retained because the pre-RNA template directs the synthesis of an RNA molecule with a complementary sequence, requires the pre-RNA to be capable of template-based replication, and be able to form a duplex with RNA. TNA and PNA, to give two examples, do form stable hybrid double-helices with RNA; p-RNA and pentuloNA do not. Of course, it is possible that RNA could have been preceded by a sequence of pre-RNAs, and needed only to form a duplex with the penultimate nucleic acid.[243] This genetic system does not have to be a nucleic acid: one theory suggests that Watson-Crick base pairing could have taken place between purines in different conformations (thereby circumventing the difficulties of prebiotic pyrimidine synthesis),[247] later modified to suggest that this all-purine system was the result of a gradual take-over of an imidazole-based system.[29] Another suggests that mixed polymers containing a variety of moieties are capable of base-pairing and catalysis and therefore are the evolutionary ancestor to both nucleic acids and proteins.[248]

An alternative transition is the discontinuous transition, or genetic take-over:[84] a pre-existing self-replicating system evolves a synthesis of RNA (possibly to fulfil a functional role), and is then taken over by RNA. Genetic information is not retained, and there is no requirement for a structural relationship for template-based replication: it has even been suggested that life evolved from inorganic crystals which were able to

store a large quantity of information, followed by a genetic take-over by organic moleucles.[249]

It is not clear why a pre-RNA system would then vanish (the transition we have to compare it with – that of RNA to DNA/protein – is not yet complete). The pre-RNA World theory would carry more weight if remnants of it were apparent in life today, even if the earlier genetic system did not maintain such a prominent role in extant metabolism as RNA.[86]

Another theory is that the other pre-RNA candidates did not invent RNA, but would have co-existed with it in a heterogeneous system[235] (a co-evolution of deoxy- and ribonucleotides, resulting in chimeric biopolymers, has also been suggested[250]). RNA would then result from continuous incremental replacement of the previous chimeric system: this avoids complete reinvention of chemistries associated with the different systems, and allows for chemical evolution through selection and refinement of the nucleic acid backbone and nucleobases based on the emergent properties of RNA. There may have been some advantages associated with a highly heterogeneous system: it has been shown that functional RNAs can tolerate the structural variety associated with a mixture of 2′,5′- and 3′,5′-linkages,[142] and it is possible that further heterogeneity of backbone sugars could also be tolerated. It is also possible that RNA is a vestigial system, or 'frozen accident': that the canonical nucleotides were more accessible to prebiotic chemistry than non-natural equivalents, and the structure that was 'selected' as a result of this has remained.[242] This argument would be more persuasive if compelling prebiotically plausible routes to all the canonical nucleotides were known. The argument that RNA is an emergent structure, with novel properties that only arise with a specific level of structural complexity when formed from components with less complexity, relies on the understanding that RNA is made up of specific components and that alternative disconnections are not available. A systems chemistry approach, based on selective and high yielding controlled reactions, may solve this problem. Until all the potential chemical routes to RNA from plausible feedstock molecules are exhausted, the assumption that constitutional self-assembly of RNA did not occur, and that another genetic system preceded the RNA world, is premature.[251]

1.4.5 A new approach to nucleotide synthesis

John Sutherland has postulated that the RNA World hypothesis is best tested through chemical synthesis: if RNA nucleotides can be synthesised under prebiotic conditions in a laboratory, it is possible that they could be synthesised on the early Earth. If their synthesis on a significant scale proves impossible when guided by a chemist then it is unlikely that it would have occurred naturally prebiotically, and therefore the RNA World is less likely to have existed.[91] As it appears that the seemingly easiest synthesis, *via* the most obvious retrosynthesis, was unlikely to occur under prebiotic conditions, focus instead has turned to other, more obscure routes. The systematic investigation of these can provide evidence for the plausibility of RNA abiogenesis.

Sutherland and co-workers hypothesised an alternative retrosynthesis of nucleotides *via* amino-oxazolines **60**, thereby bypassing both ribose *ribo*-**30** (ribose amino-oxazoline *ribo*-**60** degrades around 70 times more slowly than *ribo*-**30**[252]) and nucleobases (and therefore the formation of the glyosidic bond) as discrete entities. Orgel and Sanchez had previously showed that stepwise assembly of the base on a sugar template could be used to circumvent the problem of glycosidic bond formation (***Figure 1.24***, top).[253] Ribose *ribo*-**30** and arabinose *ara*-**30** react efficiently with cyanamide **4** to give the corresponding amino-oxazolines:[253] the significant stability of these compared to the free sugars indicates a potential solution to the stability problem of ribose in nucleotide synthesis.[252] Reaction with cyanoacetylene **5** provides anhydronucleosides **63**, which hydrolyse to give α-ribocytidine α-*ribo*-**29** (the correct sugar isomer, but the wrong anomer) or ß-arabinocytidine ß-*ara*-**29** (the desired anomer, but the incorrect sugar isomer).[253] Irradiation with UV light results in photoanomerisation, but this is very inefficient: the highest yield of the natural nucleotide ß-*ribo*-**29** that has been isolated from the irradiation of α-*ribo*-**29** is 6%.[253,254] (One of the major by-products of this reaction is oxazolidinone **64**: its formation can be blocked by using a 2′-phosphorylated derivative of α-*ribo*-**29**, resulting in an increased conversion to the desired product.[255] This is not a solution, however, as it raises another problem: the selective prebiotic 2′-phosphorylation of α-*ribo*-**29**.)

Tapiero and Nagyvary demonstrated an alternative route to cytidine nucleotides from anhydrocytidine derivatives (***Figure 1.24***, bottom).[256] β-cytidine-2′,3′-cyclic-monophosphate ß-*ribo*-**29**-2′,3′cP (which is activated towards ring-opening by P-O bond cleavage, and therefore a potential substrate for oligomerisation) can be

produced from the rearrangement of *ara*-**63**-3′P – the same anhydronucleotide (but with a 3′-phosphate group) as in the synthesis of Orgel and Sanchez – in aqueous solution at around pH 6. [256] Tapiero and Nagyvary did not manage to produce ß-*ribo*-**29**-2′,3′cP cleanly by this method, as the hydrolysis product ß-*ara*-**29**-3′P was also formed, especially in more basic solution. *Ara*-**63**-3′P can be accessed through reaction of arabinose-3-phosphate with cyanamide **4** and cyanoacetylene **5,** in the same manner as for the unphosphorylated equivalent.[257]

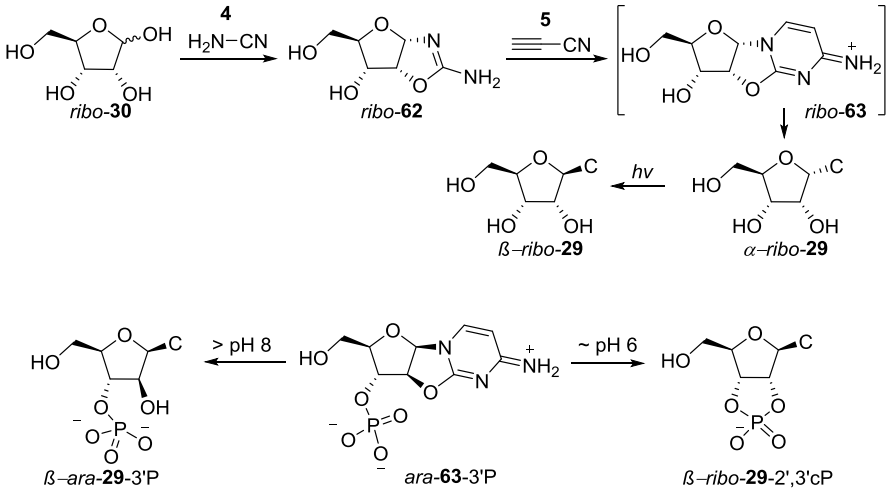


Figure 1.24*: Prebiotically plausible syntheses of cytidine* ***30*** *and phosphorylated derivatives (C = cytosine* ***31)****.* ***Top****: Stepwise synthesis of the base on a sugar scaffold, with subsequent photoanomerisation to the canonical nucleoside ß-ribo-****29****.* ***Bottom****: pH-dependent behaviour of a phosphorylated derivative.*

Aminooxazoline **62**, in a prebiotic context, could be formed from a pentose sugar, then acting as a stable store for this sugar, but the pentose sugars could be bypassed altogether: **62** can instead form from a masked aldol reaction between glyceraldehyde **33** and 2-aminooxazole **26**. It had previously been shown that 2-aminooxazole **26** can be formed in good yield from the reaction of cyanamide **4** and glycolaldehyde **32**, both of which are presumed to be present on the prebiotic Earth. This occurs, however, at a pH high enough to impair the next steps in the proposed synthesis;[258] at neutral pH in water the reaction was low-yielding due to a drop in reaction rate in the absence of a base catalyst.[138] Investigations by Sutherland and colleagues revealed the value of inorganic phosphate as neutral-pH general base catalyst, allowing a yield of 80% **26** to be achieved (***Figure 1.25***).[138]

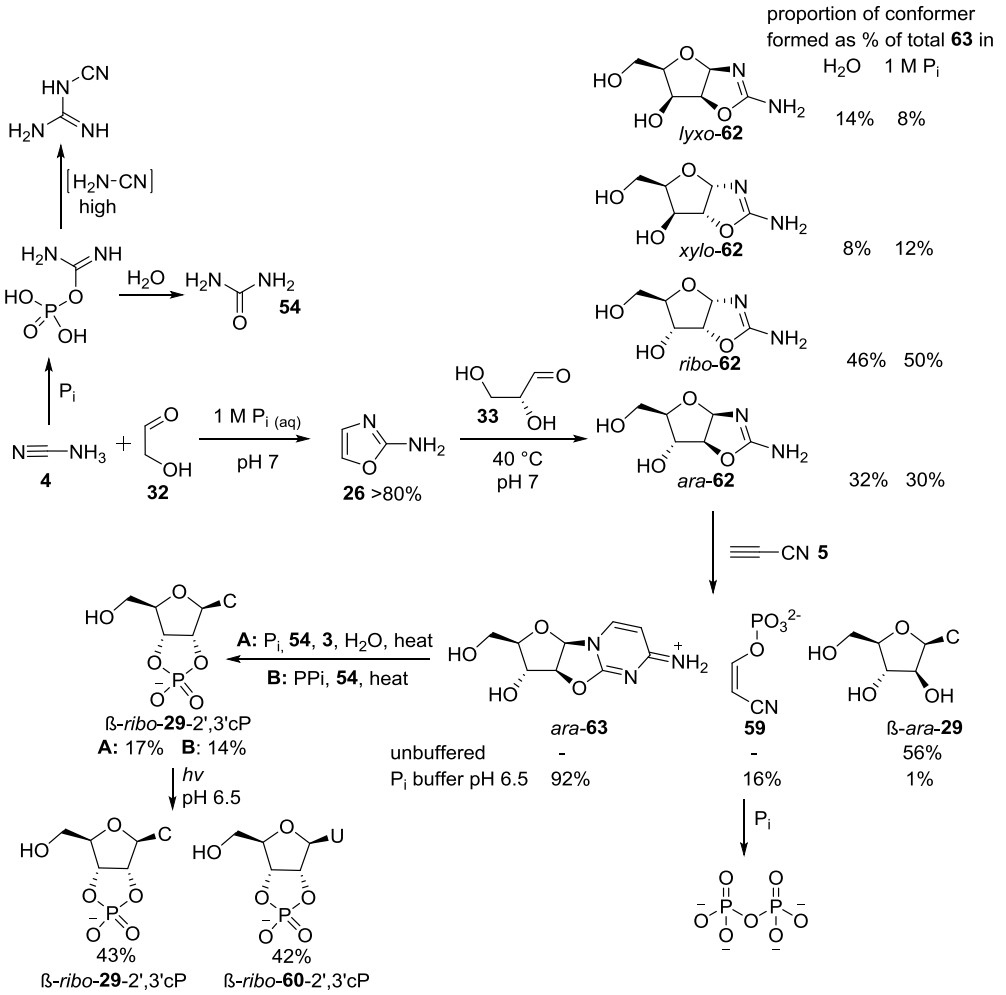


Figure 1.25*: A complete, prebiotically plausible synthesis of the pyrimidine ribonucleotides from two- and three-carbon units. (C = cytosine **31**, U = uracil **51**)*

The reaction of **26** with glyceraldehyde **33** in water at neutral pH is clean and high-yielding, with a stereochemical preference for the formation of *ribo*- and *arabino*-aminooxazolines (*ribo*-**62** and *ara*-**62**);[259] the reaction is improved in the presence of residual inorganic phosphate, with the unwanted lyxose derivative *lyxo*-**62** becoming a minor product.[138] Selective crystallisation of *ribo*-**62** (the least soluble of the pentose-amino-oxazolines) provides further enrichment of the desired *ara*-**62** in solution, in a warming and cooling cycle which potentially mimics the day-night cycle on early Earth.[138,259] The reaction of *ara*-**62** and cyanoacetylene **5** in unbuffered aqueous solution at pH 7 ultimately results in the hydrolysis of the anhydronucleoside *ara*-**63** formed, due to a rising pH during the reaction, thereby giving β-*ara*-**29**.[138] When inorganic phosphate was again used as a buffer at near-neutral pH in the reaction of **5** with *ara*-**62**, *ara*-**63** was produced cleanly with no hydrolysis.[138]

It had previously been shown that the 3′-phosphate derivative anhydronucleoside *ara*-**63**-3′P can undergo competing rearrangement, with inversion of the 2′-sterochemistry providing ß-*ribo*-**29**-2′,3′cP, and hydrolysis to ß-*ara*-**29**-3′P in unbuffered aqueous solution at neutral pH. [257] *Ara*-**63**-3′P can be regioselectively generated *in situ* from *ara*-**63** in two ways: by dry-state heating with pyrophosphate and urea **54**, or by heating with inorganic phosphate and **54** in formamide **3** solution (pyrophosphate is a byproduct of the preceding step; **54** is a byproduct of the first step of the synthesis).[138] The fact that phosphorylation predominantly occurs at the 3′-hydroxyl is perhaps surprising: usually primary alcohols are less sterically hindered and therefore should be phosphorylated faster. The regioselectivity observed is due to stereoelectronic effects: an X-ray crystal structure of the anhydronucleoside shows that its conformation holds the 5′-hydroxyl in a position that makes it unusually hindered;[138,260] this proximity also leads to the delocalisation of the O-5′ lone pair over the antibonding orbital of C-2=N-3, reducing the nucleophilicity of O-5′.[260] Intramolecular nucleophilic substitution then gives ß-*ribo*-**29**-2′,3′cP.[138]

When the products of this phosphorylation are irradiated in aqueous solution with ultraviolet light, ß-*ribo*-**29**-2′,3′cP is partially hydrolysed to the uridine derivative ß-*ribo*-**62**-2′,3′cP, with other pyrimidine nucleosides and nucleotides being destroyed.[138] This provides a useful way to purify the desired activated nucleotides. Although the 2′,3′-cyclic phosphate ring can undergo hydrolysis to the unactivated nucleotide-2′- and 3′-phosphates, cyanoacetylene **5** drives an efficient cyclisation, reforming the activated cyclic phosphate, under the same pH conditions as the cyanovinylation reaction involving **5** earlier in the synthetic route.[120]

The intermediacy of 2-aminooxazole **26** in this synthetic route, and its constitutional relationship to 2-aminoimidazole **25** (an efficient catalyst of nucleotide oligomerisation[136]) is particularly noteworthy. The two can be formed under the same conditions (***Figure 1.26***): at pH 7 in the presence of equimolar glycolaldehyde **32**, cyanamide **4,** phosphate and NH_4Cl, **26** and **25** form in equimolar quantities.[137] Increasing the concentration of NH_4Cl to 5 M results in near-exclusive formation of **25**; acidic pH also favours the formation of **25** (as the dehydration step required for **26** synthesis does not occur as readily at lower pH). When exposed to glyceraldehyde **33**, **25** (which has higher aromatic stability) is left largely unreacted, whilst **26** (which

contains a carbon atom sufficiently nucleophilic) readily reacts to provide the aminooxazolines. The common selective origin of **25** and **26** and their distinct reactivates indicates a potential prebiotic reaction network that could lead to both the synthesis of ribonucleotides and their activation to oligomerisation.

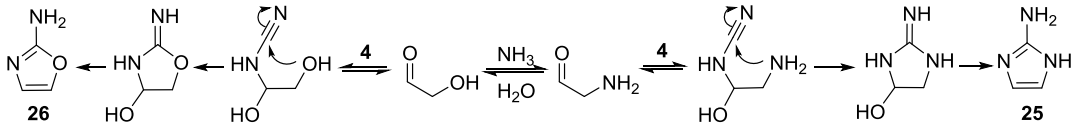


Figure 1.26**: Possible mechanism for the synthesis of **25** and **26.

Despite the attractions of this route, it does not take full advantage of one of its major by-products: ribose aminooxazoline. The high crystallinity of *ribo*-**62**, and its greater tendency to crystallise than its stereoisomers,[253] offers a prebiotically plausible means for its purification and accumulation. Furthermore, *ribo*-**62** crystallises as a conglomerate (that is, the different enantiomers form separate crystals),[259] which may offer a means for the amplification of optical asymmetry on the early Earth. Unfortunately, given this beneficial crystallinity and the high yield of *ribo*-**62** in the reaction of glyceraldehyde **33** and **26** (and an even higher yield of the phosphorylated derivative *ribo*-**62**-5′P in the reaction of glyceradehyde-3-phosphate **33**-3P and **12**[261]), the synthetic route outlined above requires *ara*-**62**. Although methods exist for converting *ribo*-**62** to the arabino isomer,[262] and for α-cytidine to the ß-anomer,[253,254] they are both inefficient and involve nucleoside(-precursor) destruction to yield the oxazolidinones **64**. The incubation of *ribo*-**62** in phosphate buffer results in its partial conversion to *ara*-**62** (***Figure 1.27***, blue box).[262] A plausible interconversion mechanism initiates with the ring-opening of *ribo*-**62** to give an iminuim species.[262] Phosphate-mediated deprotonation and reprotonation at C2′, followed by ring closure, results in the regeneration of *ribo*-**62** or in the formation of *ara*-**62**. The oxazolidinones **64** observed as by-products could either form from the direct hydrolysis of the aminooxazolines, or from indirect hydrolysis of the iminium intermediates. Oxazolidinones **64** do not react with cyanoacetylene **5,** but nevertheless their formation lowers the yield of anhydronucleoside. The highest ratio of *ara*-**62** to *ribo*-**62** (2:1) is observed at pH 5 and elevated temperature (60 °C), but these conditions also result in significant conversion to **64** (77%).[262]

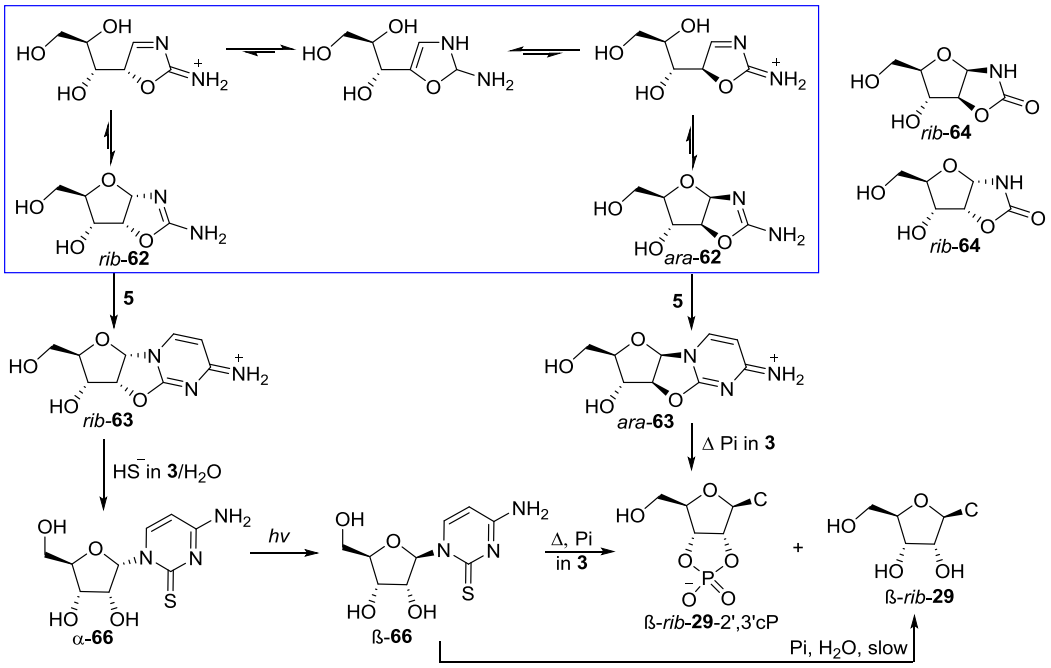


***Figure 1.27**: Synthesis of ß-cytidine nucleotides and nucleotides from ribo-**62**. Blue box: proposed mechanism for the interconversion of ribo-**62** and ara-**62** in phosphate buffer. Thiolysis of ribo-**63** gives α-**66**, which undergoes photoanomerisation to ß-**66**. Incubation of ß-**66** in phosphate buffer leads to the slow formation of ß-ribo-**29**; phosphorylation in formamide **3** leads to the formation of ß-ribo-**29** and ß-ribo-**29**-2',3'cP.*

The irradiation of a thiolysed derivative of α-cytidine, however, leads to very different results (***Figure 1.27***).[128] It has been found that thiolysis of ribo-anhydrocytidine *ribo*-**63** leads to the formation of α-2-thioribocytidine α-**66** in good yield (84%), and subsequent UV irradiation results in epimerisation to ß-**66** (76%).[128] Hydrolysis in phosphate buffer at elevated temperature resulted in the slow production (84 days) of the canonical ß-ribocytidine (41%); alternatively, the direct phosphorylation in the presence of inorganic phosphate resulted in the formation of ß-ribocytidine (48%) and its 2′,3′-cyclic phosphate derivative (24%) – phosphorylated derivatives of ß-2-thioribocytidine were not detected. Although the thiolysis step takes place in aqueous formamide **3,** and the phosphorylation step in anhydrous **3** (which could be achieved simply by heating), the intermediate photoanomerisation takes place in formamide-free aqueous solution, and it is not clear how this change of solvent could be easily achieved on the early Earth. Nevertheless, if enantiopure crystalline *ribo*-**62** could be produced, which relies on a method of producing glyceraldehyde **33** in sufficient enantiomeric excess, this could be a promising short and high-yielding route to the enantiopure canonical pyrimidine nucleotides.

1.5 A systems chemistry approach

The synthetic dead end that resulted from the separation of nitrogenous and oxygenous chemistry, under the assumption that the two would negatively interfere, in attempting prebiotically plausible construction of nucleotides has been resolved. Life itself is based on interconnected systems, with function arising from the interaction of their molecular components.[263] It therefore follows that early life forms, and the prebiotic world, also required the interaction of multiple systems and components to an even greater extent than has been demonstrated in Sutherland's nucleotide synthesis.

Contrary to the previous reductionist approach to understanding nucleotide abiogenesis, in which the separate systems of life were studied separately, a systems chemistry approach may prove advantageous.[92,263–265] An exact definition of systems chemistry is lacking, but the use of multicomponent and sequential one-pot multistep reactions are a common theme. Networks of reactions respond to stimuli in complex ways, meaning complex behaviour can emerge from networks:[264] the addition of further components may counterintuitively simplify the range of products formed, instead of complicating them further. A simple example of this is the inclusion of phosphate from the start of Sutherland's nucleotide synthesis in a multifaceted role.[138] Catalysts and molecular chaperones, either present *ab initio* or formed *in situ,* are capable of directing reactions to restrict unwanted products so that multiple species do not have to react in isolation.

Whilst Sutherland's seminal synthesis of pyrimidine ribonucleotides undoubtedly has its advantages, there is a strict requirement for the sequential addition of cyanamide **4** to glycolaldehyde **32** (forming 2-aminooxazole **26**), followed by glyceraldehyde **33**. **26** has a high vapour pressure, and so theoretically **32** and **4** could react in one location, the **26** formed sublime, then rain into a pool containing **33** at a different location. Prebiotically plausible separate syntheses of **32** and **33** are lacking, however, and to make matters more complicated, **33** interconverts to its more stable isomer, dihydroxyacetone **34**, especially in phosphate buffer.

A physiochemical mechanism involving the use of 2-aminothiazole **67** as a chemical chaperone solves this problem by enabling the selection, accumulation and time-resolved sequestration of glycolaldehyde **32** and glyceraldehyde **33** (***Figure 1.28***).[266] **67** has a higher degree of aromaticity than 2-aminooxazole **26**, preventing masked

aldol reactions with **32** and **33**. Instead, the sugars sequentially react to form an aminal which selectively crystallises: glycolaldehyde aminal **68** first crystallises over 2 h, followed by glyercaldehyde aminal **69** over 2–17 d. This large time resolution allows the separation, accumulation and subsequent reaction of the aminals. It also provides a mechanism for the selection of **33** over **34** and the accumulation of the C_2 and C_3 sugars required for nucleotide synthesis from a complex mixture of sugars. **67** itself can be formed in a similar manner to **26** and **25** from cyanamide **4** and ß-mercaptoacetaldehyde **70**.[266] The high nucleophilicity of sulfur allows the rapid and selective trapping of **70** by **5,** even in the presence of a higher concentration of α-hydroxyaldehydes.

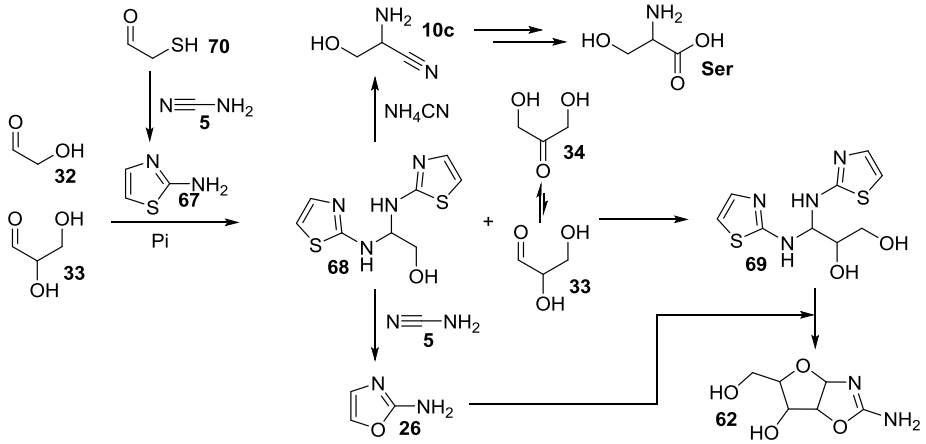


***Figure 1.28**: 2-Aminothiazole **67** –controlled aldehyde sequestration provides a prebiotic method for the selection of proteinogenic amino acids and the synthesis of nucleotides. The time-resolved crystallisation of **68** and **69** allows a method for the separation of glycolaldehyde **32** from **33**. Reaction of the crystalline aminals with ammonium cyanide provides the α-aminonitriles, such as serine aminonitrile **10c**, from a complex mixture of aldehydes and ketones.*

2-Aminothiazole **67** can play a similar role, controlling the outcome of the reaction of a complex mixture of starting materials, in proteinogenic amino acid synthesis.[266] **67** does not react with ketones, but does form stable crystalline aminals with aldehydes, indicating a possible prebiotic process for their accumulation and purification, allowing for the selective synthesis of the natural (α-H) amino acids.[266] This common molecular chaperone, operating using the same physiochemical mechanism, therefore enables the selection and accumulation of nucleotide and proteinogenic amino acid precursors, providing another link between nucleic acid and peptide chemistry.

Investigation by the Sutherland group has revealed plausibly prebiotic conditions under which the sugar precursors for nucleotide synthesis can be produced,[267,268] alongside proteinogenic amino acid precursors, and the hydrophilic moiety of the lipid bilayer from hydrogen cyanide.[64] The generation of nucleic acids, peptides and protocell membranes are therefore linked from the formation of their building blocks. Under the same conditions, ribose is reduced to 2-deoxyribose, and a uracil-derivative to thymine, implying that the biosynthetic precursors of DNA can be abiotically produced from those of RNA.[269]

Glycolaldehyde **32** and glyceraldehyde **33** can be produced from the copper-catalysed photoreduction of glycolonitrile (the unwanted product of the combination of the formose reaction and nucleobase synthesis) using hydrogen sulfide in aqueous solution, with phosphate acting as a pH buffer.[268] In fact, this synthesis can be initiated with the photochemical Kiliani-Fischer homologation of hydrogen cyanide, with hydrogen cyanide as the reductant as well as the starting material.[267] Photoredox cycling of cyanocuprates oxidises two molecules of hydrogen cyanide to give cyanogen **6** (subsequently hydrolysed to cyanate **49**, which unfortunately can trap the desired sugar products as cyclic species), with the release of reducing power in the form of protons and hydrated electrons.[267] These can be used to reduce a further molecule of hydrogen cyanide to, after hydrolysis, formaldehyde **1** (ammonia is therefore released, contributing to a prebiotic source of reduced nitrogen). The reaction of **1** with hydrogen cyanide forms glycolonitrile **72a** (thereby fixing carbon as a non-volatile and stable form), and iteration of the reduction process results in the sequential formation of glycolaldehyde **32**, its cyanohydrin, and glyceraldehyde **33**. This exploits the nucleophilicity of the cyanide ion so that homologation occurs through the formation and selective reduction of cyanohydrins, circumventing the umpolung reactivity required in the dimerization of formaldehyde **1** to give **32**. Glycolonitrile **72a** has been postulated as a more plausible starting material for this synthesis than hydrogen cyanide: **72a** could be produced from the reaction of hydrogen cyanide (rained into surface water bodies from the atmosphere where it could be produced by impact shock) and **1** (also rained in, from the photoreduction of carbon dioxide in the atmosphere).[268] This synthesis has a further point of interest: formation of glyceraldehyde **33** *via* sequential cyanide capture and photochemical reaction bears a similarity to the sequential enzyme-catalysed carbon dioxide capture and reduction of the Calvin cycle, which produces glyceraldehyde-3-phosphate **33**-3P.[119]

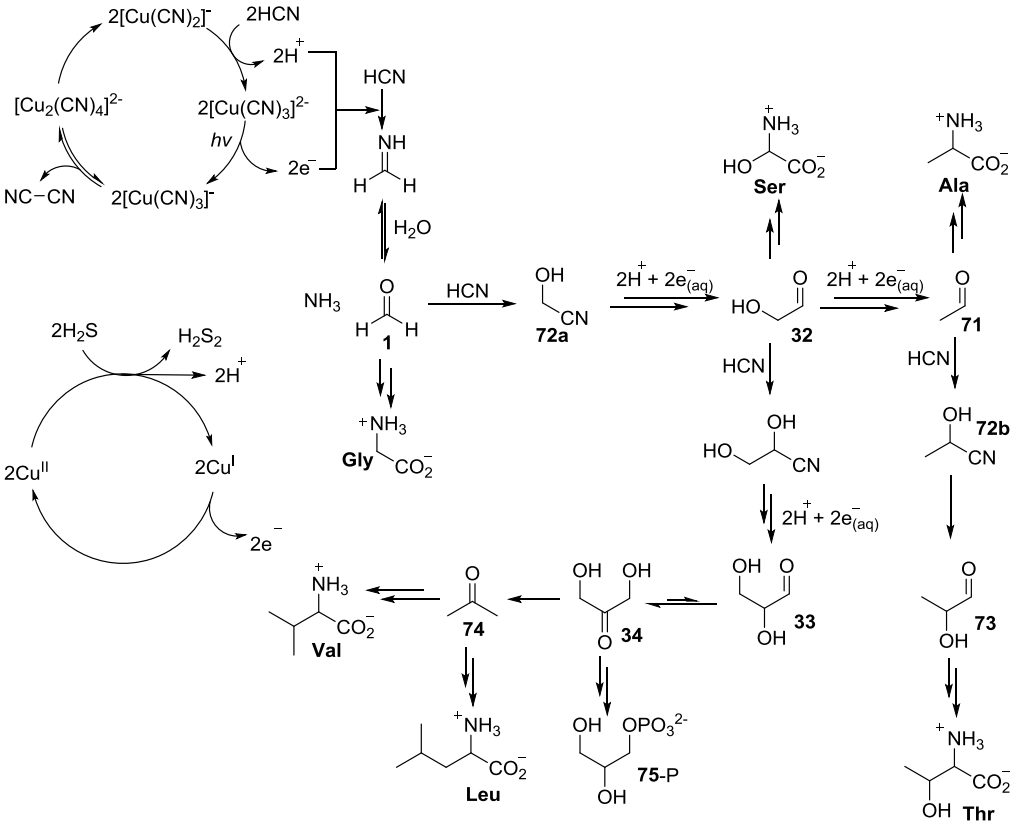


Figure 1.29*: Hydrogen-cyanide-cyanocuprate systems photochemistry and photoredox systems chemistry with hydrosulfide as the ultimate reductant, giving rise to a number of nucleotide, amino acid and lipid precursors.*

The sugar species produced are important in the context of ribonucleotide abiogenesis, but the production of acetaldehyde **71** (from the photochemical α-deoxygenation of glycolaldehyde **32**) and formaldehyde **1** is also significant. These are precursors of amino acids, suggesting a concurrent prebiotic synthesis of ribonucleotides and amino acids. Addition of yet another molecule of hydrogen cyanide and ammonia to an aldehyde generates the corresponding aminonitrile[268] (although the aminonitrile can also be formed from the addition of ammonia to the corresponding cyanohydrin, which dissociates to cyanide and aldehyde[64]), followed by hydrolysis to the amino acid. In this manner, glycine **Gly** can be produced from formaldehyde **1**, alanine **Ala** from acetaldehyde **71**, and serine **Ser** from glycolaldehyde **32**. Lactonitrile **72b**, the cyanohydrin of acetaldehyde, can undergo reductive homologation to produce lactaldehyde **73**, a threonine **Thr** precursor. The phosphate-catalysed isomerisation of glyceraldehyde **33** to dihydroxyacetone **34**, followed by two sequential α-deoxygenations, provides acetone **74**, which provides a point of access to the branched amino acids valine **Val** and leucine **Leu** contributing to a prebiotic source of

reduced nitrogen *via* further reductive homologations.[64] The copper-mediated cross-coupling of cyanide and acetylene produces acrylonitrile, and ß-aminopropionitrile upon reaction with ammonia: a precursor of proline, lysine and arginine.[64]

The reduction of dihydroxyacetone **34** also produces glycerol **75**.[64] Phosphorylation (under the conditions used for nucleotide synthesis[138]) yields glycerol phosphate **75**-P,[64] which, along with fatty acid, is one of the components of phospholipids.[57] A link between a genetic system and a compartment-forming system could therefore be established from the synthesis of their respective precursors.

1.6 Outlook

Whilst the RNA World theory is certainly plausible – the ubiquity of RNA in modern life and its dual functionality as genotype and phenotype, allowing an information flow, make RNA a strong contender for the first (informational) biomolecule of life – it is unlikely that RNA would have acted in isolation on the early Earth. The integration of genetic information and proteins in today's life indicate the importance of a prebiotic synthesis of amino acids and peptides.

The results obtained in recent years from taking a systems chemistry approach to studying the origin of life show the advantage of taking a comprehensive and integrated view. Any further investigations of the chemical origins of the integrated informational, catalytic and compartment-forming systems that make up life should preferably take place under the conditions established for other prebiotically plausible networks, and using common reagents. Nevertheless, it is still not evident what conditions and reagents would have been present on the abiotic Earth: therefore, all feasible routes towards the biomolecules necessary for life should be explored to refine our knowledge of prebiotic chemistry. Determining where reactivity is not observed helps us understand what plausible chemical scenarios were available on the early Earth.

The work presented here covers investigations towards the synthesis of peptides and nucleotides. We demonstrate a selective synthesis of amino acid precursors, equipped with synthetic handles for possible oligomerisation, at neutral pH; and also present work undertaken with the view of exploring alternative routes towards nucleotide assembly.

2. The phosphoro-Strecker reaction: a route to amino acid derivatives

2.1 Introduction

Prebiotic chemistry seeks to explore not only the potentially prebiotic syntheses of important biomolecules, but also understand, from a systems chemistry perspective, how the substructures of these molecules are related. This approach can lead to an understanding of not only how life emerged, but also why the molecules essential to life were selected, and how these then evolved. Proteins and nucleic acids are fundamentally linked in extant life through the central dogma of molecular biology. The RNA World hypothesis does not preclude the presence of amino acids and peptides on the early Earth;[56] indeed, it has been suggested that peptides played an essential role in an RNA-based life, for example, by enhancing ribozyme function.[68] Showing that the formation and activation of peptides and ribonucleotides could be linked on the early Earth would lead to further insights into their selection and evolution.

Seminal work by Miller demonstrated that amino acids **7** could be formed from sparking a mixture of methane, ammonia, water vapour and hydrogen, mimicking the introduction of energy (in a form such as lightening) into a model of the (supposed) early Earth atmosphere.[18] It seems likely that the amino acids detected are produced through the Strecker reaction from the aldehydes **8** and hydrogen cyanide formed *in situ*. Reaction of an aldehyde, **8**, with ammonia (present in the atmospheric model used in Miller's experiment) forms an imine **9,** which upon nucleophilic attack of cyanide produces α-aminonitrile **10** (***Figure 2.1***, orange box). This reaction is reversible: α-aminonitriles decompose in solutions sufficiently acidic to convert NH_3 to NH_4^+, and in solutions sufficiently alkaline to convert HCN into CN^-.[72] Therefore, the pH most favourable for the formation of aminonitriles over cyanohydrins **72** (formed from the reaction of HCN and aldehydes, ***Figure 2.1***, green box) is around pH 9.2.[270] Hydrolysis of aminonitriles produces the corresponding amino acids (in Miller's experiment, this likely occurred during the purification procedure which used strong acid[18]).

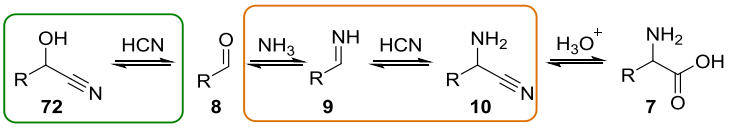


Figure 2.1: *Prebiotic synthesis of amino acids through the Strecker reaction. Sequential reaction of* ***8*** *with ammonia and cyanide forms* ***10*** *(orange box), followed by hydrolysis to* ***7****. The formation of cyanohydrins (****72****, green box) is also shown.*

Retrosynthetic analysis of the Strecker reaction allows identification of the aldehyde precursors required to access the canonical amino acids, although the aldehydes **8** can be present in the form of the corresponding cyanohydrins **72.** The reductive homologation of hydrogen cyanide in Sutherland's cyanosulfidic reaction network yields aldehydes **8** which are trapped as cyanohydrins **72**. Addition of ammonia provides the corresponding aminonitrile **10** followed by hydrolysis to the amino acid **7**.[64] Sutherland's cyanosulfidic network yields precursors common to nucleotides, amino acids and lipids, indicating a common cooperative origin for these systems.

2.1.1 Amino acid condensation

An amide bond is formally formed through the condensation of a carboxylic acid and an amine with the release of a water molecule. Therefore, the most obvious route to peptide formation is by condensing the amine moiety of one amino acid with the carboxylic acid moiety of another, but the large pK_a disparity between the reacting moieties of the amino acid (pK_a ~4 for COOH and ~10.5 for RNH_3^+) means that they are mutually unreactive in their native states at any pH: instead, salt formation can occur.[271] The two moieties are zwitterionic at neutral pH, present as carboxylate and amine under more basic conditions, and as ammonium and carboxylic acid under more acid conditions, but these combinations are not disposed to react. Condensation can occur, but slowly, and giving low yields;[71] the ligation of peptides is theoretically more efficient because the pK_a values of protected amino acids are different compared to the free amino acids.[272]

This lack of reactivity indicates that an activating agent is required so that polymerisation to give peptides can occur. A number of reagents used in conventional peptide synthesis activate the carboxylic acid moiety, with subsequent attack of an amino acid amine on the new electrophilic centre.[273] Indeed, this mode of activation is used in the cell. Biological peptide synthesis begins with activation of α-amino acids by phosphorylation: the enzymatic reaction of ATP with the amino acid results in the formation of an acyl phosphate, aminoacyl-adenylate **76** (***Figure 2.2***).[274] An aminoacyl-tRNA synthetase (a protein enzyme) catalyses the transfer ('charging') of the amino acid to the 2′- or 3′-OH of the terminal adenosine in a specific tRNA. The presence of this high energy ester bond in the newly formed aminoacyl-tRNA activates the charged amino acid towards attack by the free amine of another aminoacyl-tRNA, leading to the formation of a growing peptide in the ribosome.[274] In the absence of such

sophisticated enzymatic and ribosomal control on the early Earth, a chemoselective activation of peptide precursor to allow linear peptide bond formation in water must be demonstrated.

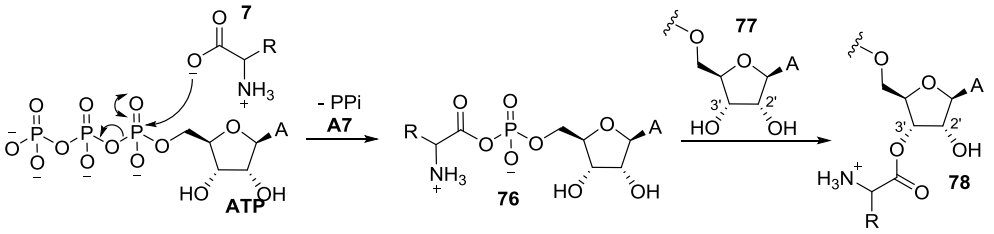


Figure 2.2: *Enzymatically controlled activation of amino acids. Coupling of an amino acid **7** to **ATP** (A = adenine) forms an aminoacyl-adenylate **76,** driven by the release of pyrophosphate. An aminoacyl-tRNA synthetase then couples the specific amino acid to the 2'- or 3'-OH of the terminal adenosine in the corresponding tRNA **77**, forming an activated aminoacyl tRNA, **78**.*

2.1.2 Prebiotic amino acid activation by phosphorylation

The centrality of phosphorylation to life, and especially the role of phosphate in extant amino acid activation, implies that phosphate could have played an important role in prebiotic amino acid activation. Rabinowitz showed that reaction of a simple amino acid (such as glycine **Gly** or alanine **Ala**) with cyclic trimetaphosphate **58** leads to the formation of a cyclic intermediate which is activated towards polymerisation: condensation between amino acids thus occurs giving dipeptides **79** (***Figure 2.3***).[74,75] (Rabinowitz indicated that a mixture of trimetaphosphate and sodium cyanide is a more effective condensing agent for amino acids – which suggests that efficient polymerisation could occur on amino acids formed *in situ* from the Strecker reaction – but did not elaborate these ideas further.[275])

Under the slightly alkaline conditions employed, it is likely that the (unprotonated) amine of the amino acid **7** attacks the phosphorus atom of cyclic trimetaphoshate **58**, forming an open chain phosphoramidate **80** (it was initially suggested that the negatively charged carboxylate attacks the negatively charged cyclic phosphate,[75] but this is unlikely). **80** can undergo ring closure with intramolecular displacement of pyrophosphate by the neighbouring carboxylate group, affording a reactive cyclic phosphoramide **18**.[276]

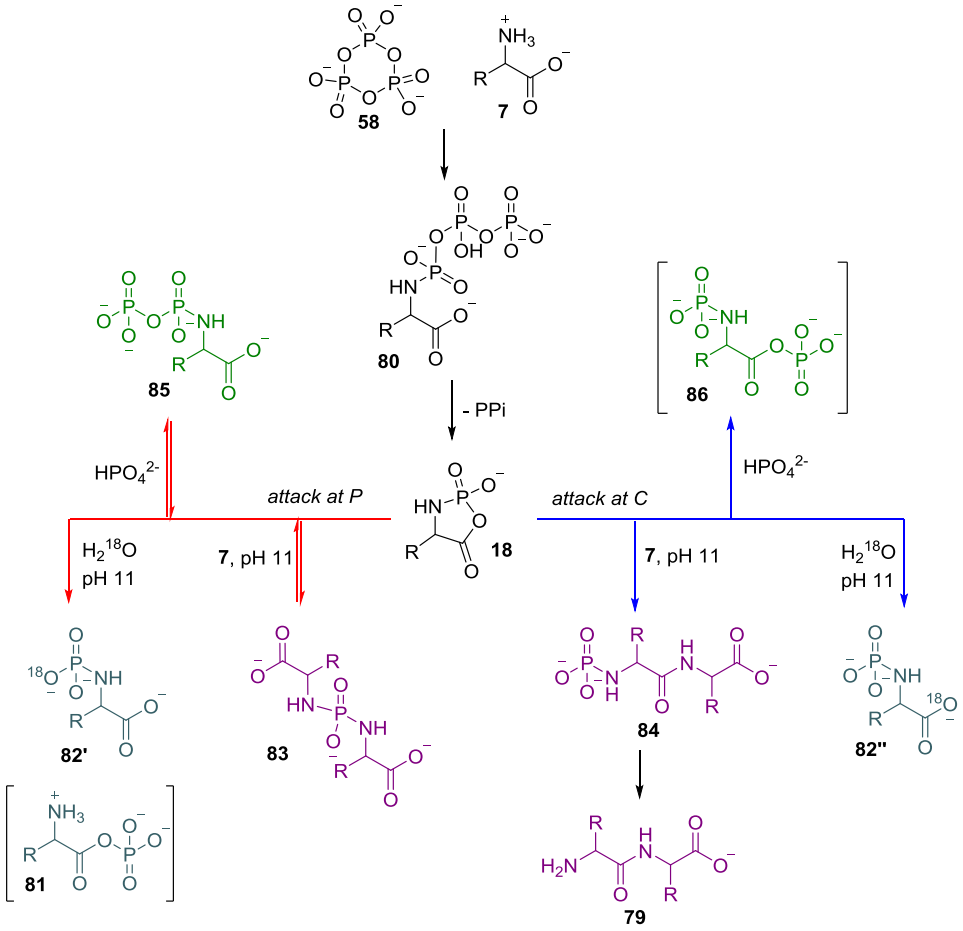


Figure 2.3: *Formation of **18** from amino acids **7** and cyclic trimetaphosphate **58** (top), and their bielectrophilicity towards $H_2{}^{18}O$ (teal products: **82**), amino acids (purple products: **83** and **84**) and phosphate (green products: **85**). Products derived from attack at phosphorous are shown via red arrows; those derived from attack at carbon are shown via blue arrows. Unobserved products are shown in brackets.*

The unstable cyclic acylphosphoramidate **18** formed contains two electrophilic sites (the phosphorus and carbonyl carbon atoms) and can react in a number of different ways, elucidated by a later study into its reactivity.[76] Rabinowitz reported that attack by water would provide the corresponding aminoacyl phosphate **81** and *N*-phosphoramidate amino acid **82**.[276,277] The aminoacyl phosphate by-product could itself be used as an activated reagent for amide bond formation.[278] However, experimental evidence revealed only the formation of the *N*-phosphoramidate amino acid **82** in alkaline aqueous conditions (***Figure 2.3***, teal): ^{18}O-isotope analysis revealed that water could attack at either electrophilic centre, although attack at phosphorous dominated in a ratio of 5:1.[76] In both cases, only the corresponding **82** was produced (albeit with the ^{18}O label on either the phosphoryl or carboxyl group). This reaction of amino acids with

cyclic trimetaphosphate and subsequent basic hydrolysis has been optimised to provide a range of 13 *N*-phosphoramidate amino acids in 60-91% yield.[279]

Nucleophilic attack by the amine group of another amino acid could also occur at either electrophilic centre (***Figure 2.3***, purple). The kinetically favoured product (the *N,N*-phosphorylated-bisamino acid **83**) is formed by attack at the phosphorus, but this product is thermodynamically unstable and spontaneously reverts back to **18**.[76] Amino acid attack at the carbonyl provides an *N*-phospho-dipeptide **84**, which would hydrolyse to the dipeptide **79**.[76,277] It is known that aminoacyl phosphates are activated towards peptide bond formation,[280,281] so it appears that the same moiety provides activation in **18** with the added benefit of ring strain. As **84** is thermodynamically favoured, the bimodal electrophilicity of **18** therefore does not reduce its potential for peptide formation. Dipeptides do not form when ß-amino acids, which are incapable of forming **18**, are used in place of α-amino acids.[282]

The cyclic acylphosphoramidate **18** can also be attacked by inorganic phosphate, forming the corresponding *N*-pyrophosphoamino acids **85** (***Figure 2.3***, green).[76] *N*-pyrophosphoglycine **85a** hydrolysed at neutral pH to give *N*-phosphoglycine **82a** and ortho phosphate: it is believed protonation of the ß-phosphate enables the carboxyl group to expel the orthophosphate anion and reform the cyclic acylphosphoramidate **18**, which is subsequently hydrolysed to afford **82**.[76] The *N*-pyrophospho derivatives of alanine (**85b**), serine (**85c**) and other amino acids were much less stable, possible because the carboxyl and phosphoryl groups are held in close proximity(due to the steric compression provided by the amino acid side chains), thereby enhancing their intramolecular carboxyl-catalysed decomposition.[76] Products derived from attack of phosphate at the carbonyl centre (**86**) are not observed, presumably due to their instability in basic aqueous conditions.[76]

Although the general mechanism of this reaction is not disputed, experimental evidence for the formation of the five-membered cyclic anhydride **18** intermediate is lacking.[76] Attempts to isolate the cyclic anhydride have been unsuccessful, due to its instability.[283] It is expected that five-membered cyclic phosphates would exhibit downfield shifted resonances in ^{31}P NMR spectra relative to the corresponding monophosphates, due to the decreased O-P-O bond angle.[283–285] Inoue and co-workers claim to have detected the presence of the cyclic anhydride (**18a**) by ^{31}P NMR

at δ ~13.6 in the reaction of cyclic trimetaphosphate and glycine,[283] on the basis that this is the expected chemical shift, although inspection of the sources cited[284,285] indicates that δ 13.6 is at the extreme upfield end of the expected range of five-membered cyclic anhydride. Inoue's assignment is disputed by Ni and colleagues who deduce this peak is *N,N'*-phosphorylated bisglycine **83a**. They report that peak is transformed into a quintuplet ($J_{H,P}$ = 8.0 Hz) in the ^{1}H-coupled ^{31}P NMR spectrum, indicating the presence of two glycine moieties with four equivalent protons; when ^{15}N-labelled glycine is used, the peak is present as a triplet in the ^{1}H-decoupled ^{31}P NMR spectrum, indicating two P-N bonds are present.[76] It is possible that such cyclic anhydride intermediates are too labile to be observed in the presence of nucleophiles such as hydroxide and amino acids.

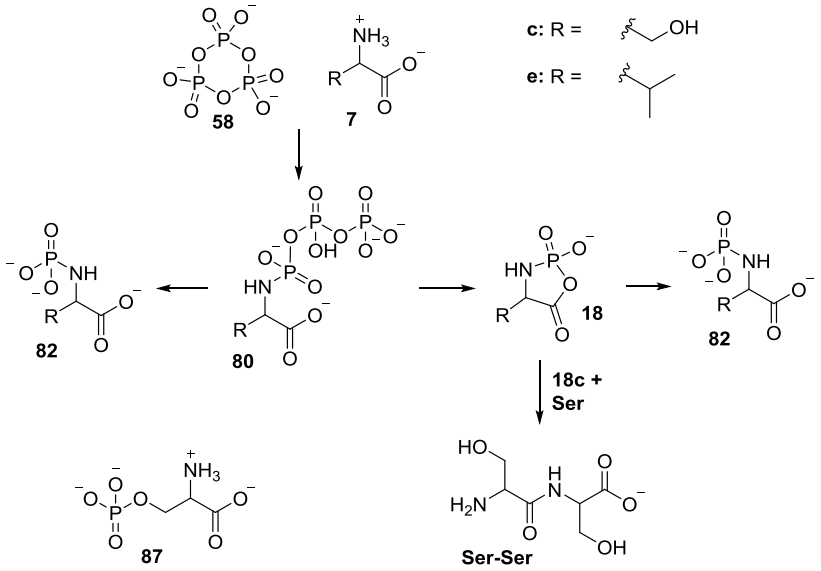


Figure 2.4: *Reaction of **Ser** (ie **7c**) and **Val** (ie **7e**) with **58**. Rabinowitz reported the formation of small amounts of the dipeptide **Ser-Ser** and O-phosphoserine **87** in the reaction of **Ser** and **58** (reaction conditions not stated but probably pH 7-8 as this is Rabinowitz's optimum pH for the reaction of glycine).*[74] *Inoue conversely reported that neither **Ser** nor **Val** formed dipeptides when reacted with **58**; instead, only the N-phosphoramidate amino acid derivatives **82** were observed (again, reaction conditions not stated but likely to be pH 12 - Inoue's optimum pH for the reaction of glycine).*[283] *The presence of pyrophosphate in the reaction mixture implies that the open chain phosphoramidate **72** of both these derivatives formed.*

The synthetic utility of this reaction in forming a range of dipeptides is unclear. Rabinowitz found that serine **7c** also provided *O*-phosphoryl derivatives such as *O*-phosphoserine **87** (***Figure 2.4***) as well as the dipeptide **79b** (only formed in small amounts).[74] Later attempts to react serine **7c** and valine **7d** with cyclic

trimetaphosphate **71** did not result in the formation of any dipeptides; instead, the *N*-phosphoramidate amino acid derivatives **82** were observed.[283] Pyrophosphate was detected as a by-product in these reactions, suggesting that the triphosphate derivatives of serine and valine (**72c** and **72d**) formed, but were very unstable and immediately hydrolysed.[283]

A number of reasons have been provided for doubting the prebiotic plausibility of this reaction as a method for the formation of peptides on the early Earth: specifically, a prebiotic source of cyclic trimetaphosphate, which furthermore must be added in large quantities over an extended period of time to allow the formation of longer oligopeptides, overcoming blocking by phosphorylation of the amino group.[277] On the other hand, it has been shown that cyclic trimetaphosphate can be regenerated from the pyrophosphate by-product of the reaction by heating at pH 4,[279] thereby offering a route to recycling this reagent on the early Earth.

Nevertheless, this activation by the formation of a cyclic phosphoramide is intriguing because it is analogous to the synthesis of nucleotides activated towards polymerisation by virtue of the ring strain imposed by a cyclic phosphate.[138] Ribonucleotides, peptides, and other molecules important to the establishment of life, may be closely linked through their mechanism of activation, and the reagents used for this activation.

An activation of simple amino acids with DAP, an ammonolysis product of cyclic trimetaphosphate, with concomitant synthesis of oligopeptides up to an octamer over a period of 40 days has recently been published (***Figure 2.5***).[208] The mechanism of activation is currently unclear, but could proceed *via* the amidophosphoimidazolide. Analysis of ^{31}P NMR spectra indicates phosphorylation of both the α-amino and α-carboxyl groups of the amino acid, with **88a**, **89a** and **82a** all being detected by NMR in the reaction of **Gly** and DAP.[208] It is known that aminoacyl phosphate esters react with amino acid esters to form dipeptides in water;[278] it is likely that the acyl-phosphoramidate **88** is the active intermediate that leads to amide bond formation when DAP is used.

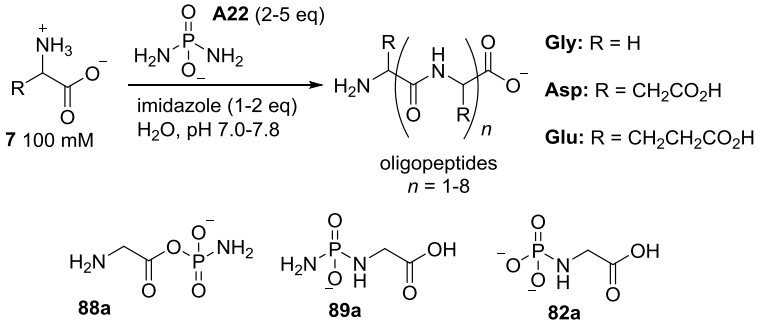


Figure 2.5: *Reaction of **Gly**, **Asp** and **Glu** with DAP in water. Oligopeptides are detected up to an octomer are detected in the reaction of **Gly,** and tetramers in the reactions of **Asp** and **Glu**. **88a**, **89a** and **82a** were detected in the reaction mixture when **Gly** was used.*

DAP has proven to be an effective phosphorylation reagent in a number of prebiotically plausible reactions on a range of substrates.[153,154,208,286] It is likely that the earliest life forms organised their metabolism around the molecules already available in the local environment, with the networks established resulting from purely chemical transformations. Interconversions between the feedstock molecules would have been refined and adapted as enzymatic control developed. The synthetic utility of DAP in a variety of potential prebiotic networks therefore indicates its potential importance on the early Earth, and so a discussion of its use as a prebiotic phosphorylating agent is provided here.

2.1.3 Prebiotic chemistry of DAP

The ammonolysis of cyclic trimetaphosphate **58** leads to the formation of amidotriphosphate (AmTP, **38**)[206,207] and, ultimately, DAP[207] (***Figure 2.6***, top). Both have been posited as potentially prebiotic reagents for the selective phosphorylation of α-hydroxyaldehydes **90** (***Figure 2.6***, bottom).[152,153] The amine functionality of these amidophosphates enables the phosphorylation of α-hydroxy aldehydes under mild conditions through intramolecular phosphate transfer *via* a cyclic phosphoramidate **61**.[152,153] Nucleophilic attack of the amine of the amidophosphate on the aldehyde moiety, forming a hemi-aminal **91**, initially tethers the phosphate at this position. The phosphate is attacked by the adjacent alcohol, with loss of pyrophosphate (if **38** is used) or ammonia (if DAP is used), forming a five-membered cyclic amido phosphate intermediate **61**. This undergoes hydrolysis with loss of ammonia to release aldehyde-2-phosphate **90**-2P.

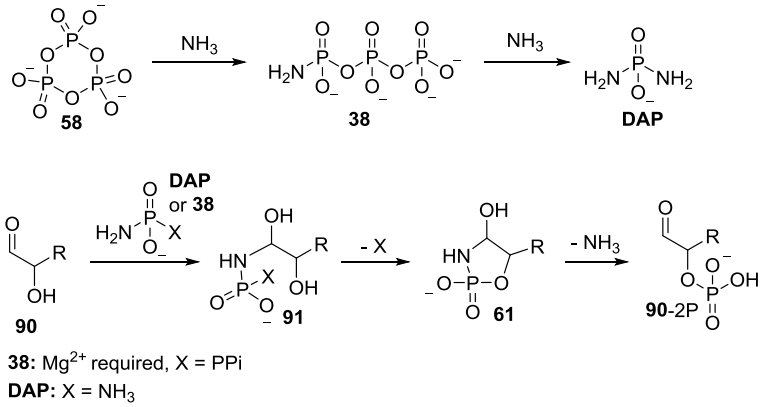


***Figure 2.6**: Ammonolysis of **58** (top), and generalised mechanism for the reaction of the ammonolysis products **38** and DAP with α-hydroxy aldehydes (bottom).*

A stoichiometric amount of Mg^{2+} is required for this reaction to proceed when AmTP **38** is the phosphorylating reagent. The exact role of Mg^{2+} has not been specifically investigated, but it most likely acts as an electrophilic activator and Lewis acid by binding pyrophosphate, increasing its lability and promoting its displacement by intramolecular nucleophilic attack, and organising the reaction centres to allow efficient substitution.[152] Mg^{2+} chelation may also assist in the hydrolysis of the transient aldehyde amidophosphate to the aldehyde-2-phosphate, and slow the reversion of the initial aldehyde-amidophosphate intermediate to cyclic triphosphate by holding the triphosphate chain in a fixed conformation.[152] Mg^{2+} is not required when DAP is used as the phosphorylating reagent,[153] increasing the compatibility with other chemistries (such as fatty acid bilayer assembly) and flexibility of conditions under which this amidophosphate phosphorylation can take place.

α-Phosphorylation prevents glycolaldehyde **32** and glyceraldehyde **33** from being used in (canonical) ribonucleotide synthesis: reaction of **32** with cyanamide **4** (coincidentally catalysed by orthophosphate) produces the sugar-nucleobase synthon 2-aminooxazole **26**, which reacts with **33** to form the stable intermediate aminooxazoline **60** (***Figure 2.7***).[138] Cyclisation to form **26** cannot occur when the C2-hydroxyl of glycolaldehyde is blocked by phosphorylation: similarly, **60** cannot form from the reaction of aminooxazole **26** with glyceraldehyde-2-phosphate **33**-2P because phosphorylation of (what would become) O-4 prevents cyclisation. The glycolaldehyde- and glyceraldehyde-phosphates **32**-P and **33**-2P are able to undergo heteroaldol addition to provide pentose-2,4-bisphosphates **41**, but this phosphorylation pattern results in a non-natural pyranosyl sugars, non-natural diphosphates and therefore non-natural

nucleosides.[287] The C-4 phosphate group of **41** prevents isomerisation to the furanose form, and the C-2 phosphate prevents the formation of aminooxazoline by reaction with cyanamide.

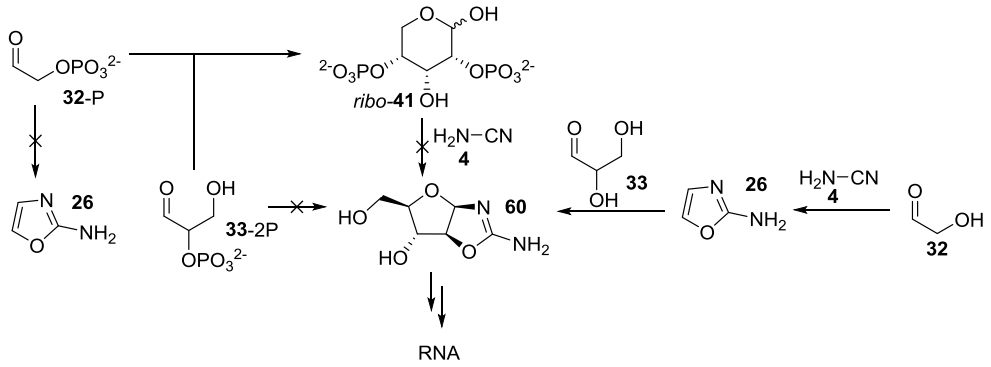


Figure 2.7: *Synthesis of pyrimidine nucleotide precursor **60** from non-phosphorylated aldoses glycoladehyde **32** and glyceraldehyde **33,** via formation of 2-aminooxazole **26**. These reactions are not possible if α-phosphorylation of **32** or **33** occurs.*

Eschenmoser hypothesised that this switch in reactivity caused by phosphorylation of sugar-precursors could mean that the prebiotic world was predisposed to the synthesis of non-canonical nucleotides;[210,287] Powner conversely hypothesised that α-phosphorylation would redirect the phosphorylated simple sugars into other metabolic pathways.[154] This would link the precursors required for two distinct proto-metabolic pathways, thereby also linking the pathways themselves.

α-Phosphorylation of simple α-hydroxyl aldehydes has also led to a prebiotically plausible synthesis of the high energy metabolite phosphoenol pyruvate **92**, which is a key intermediate in the proposed prebiotic triose metabolic network (***Figure 2.8***).[154] The formation of this metabolite is governed by the selective regiospecific phosphorylation of C2 and C3 sugars, using DAP, with the retention of the aldehyde functionality, giving regiospecific access to 2-phosphates.

Prebiotic entry to glycolysis thus occurs through a regiochemical α-phosphorylation isomer of the canonical form (extant metabolism uses glyceraldehyde-3-phosphate)[154] echoing the prebiotically plausible nucleotide synthesis reported by Powner and colleagues.[138] This potentially prebiotic ribonucleotide synthesis occurs provides a 2′,3′-cyclic phosphate, rather than terminal 5′-phosphates;[138] similarly, triose glycolysis is accessed through glyceraldehyde-2-phosphate **33**-2P (which is formed *via* a cyclic phosphate moiety) rather than the terminal phosphate isomer (glyceraldehyde-3-

phosphate **33**-3P).[154] The similarity of these model reaction conditions and the regiochemical switch from α-phosphorylation suggest that an integrated and unified system to simultaneously establish nucleotide synthesis and triose metabolism may have been important in establishing the chemical networks required for life.

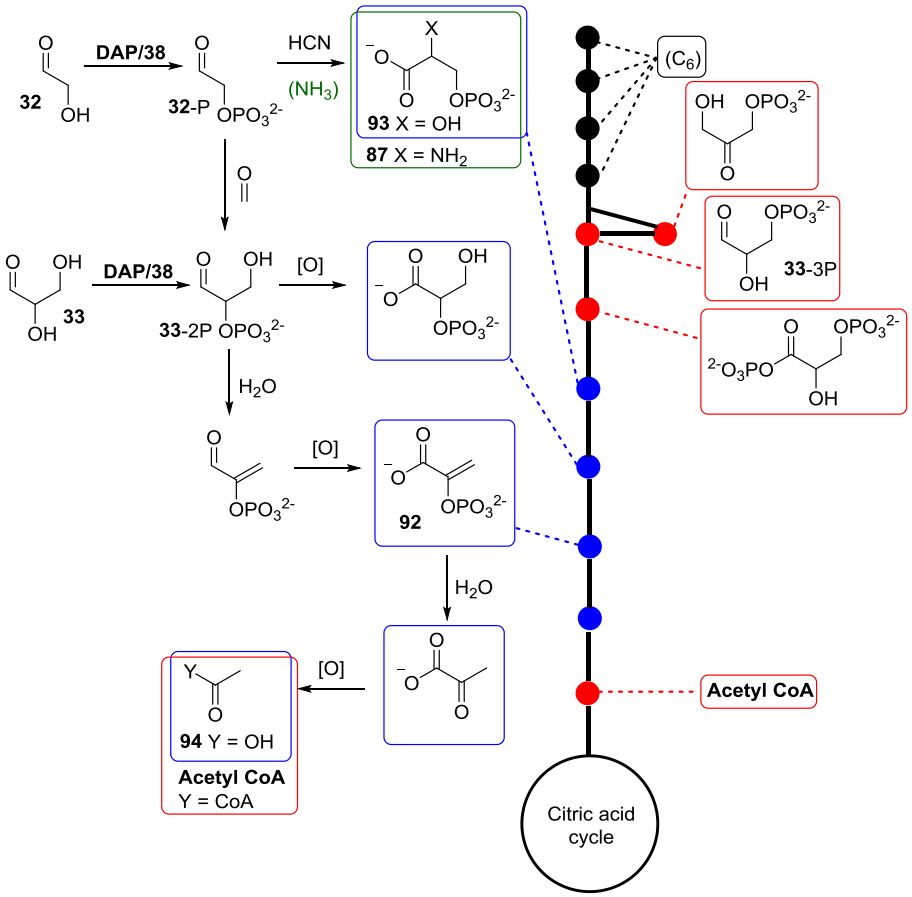


Figure 2.8: *A comparison of the proposed prebiotic triose glycolysis network (left) with the extant enzymatically controlled pathway from C_6 metabolites to the citric acid cycle (right). Metabolites in red boxes are present in modern glycolysis but are not accessed by the prebiotic network; metabolites in blue boxes are accessed through prebiotically plausible aqueous chemistry.* ***94*** *provides a tentative link to the citric acid cycle; the divergent synthesis of glyceric acid 2-phosphate* ***93*** *and O-phosphoserine* ***87*** *(green box) links glycolysis to serine metabolism.*

As well as the phosphorylation of simple α-hydroxyaldehydes such as glycolaldehyde **32** and glyceraldehyde **33**, the mechanism of this reaction also allows the regiospecific phosphorylation of the 2-OH position of aldosugars (***Figure 2.9a***).[153] The reactivity of the various aldoses is related to the relative stabilities of their open aldehyde forms in aqueous solution. The amido cyclic phosphate intermediate **95** first converts to a cyclic phosphate (the mechanism of this transformation is unclear) before hydrolysis to give the final stable 2-phosphate products. Unfortunately, the use of DAP to selectively

phosphorylate arabinose to give arabinose-3-phosphate followed by elaboration of the nucleobase has been unsuccessful.[286]

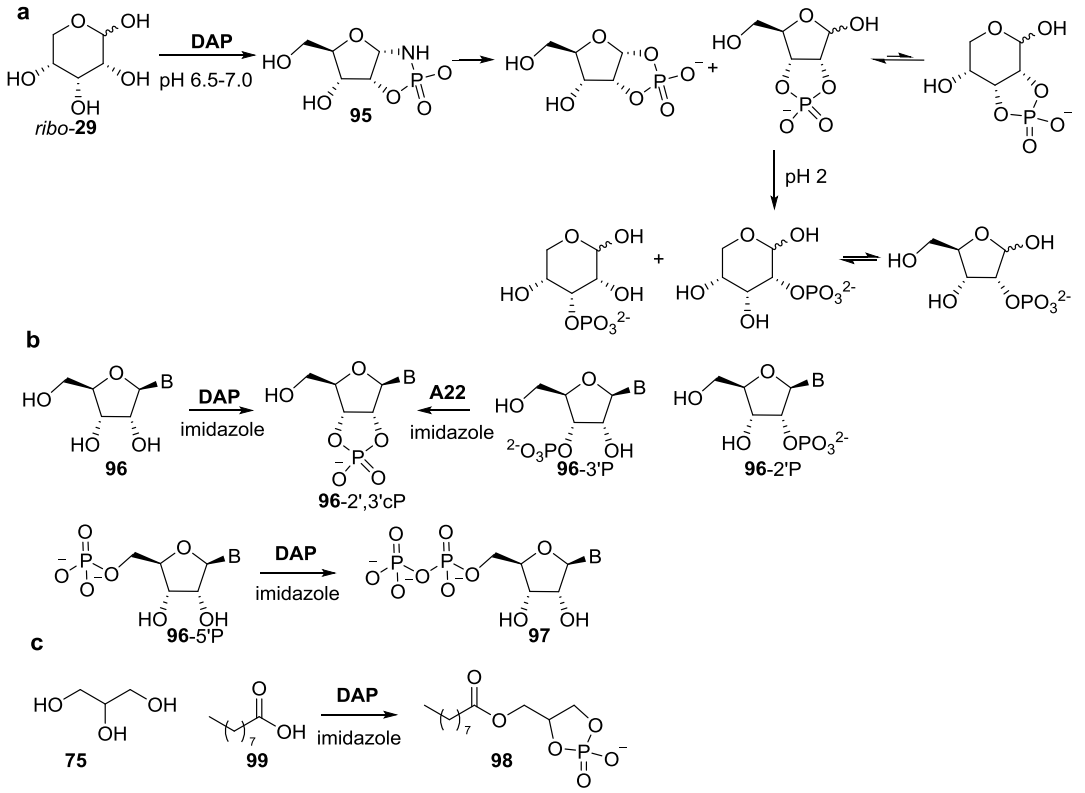


Figure 2.9*: An assortment of prebiotically plausible phosphorylations using DAP.*

a*: Reaction of ribose with DAP yields a variety of phosphorylated isomers*

b*: Phosphorylation of nucleosides with DAP yields the 2',3'-cyclic phosphate nucleotides **96**-2',3'cP; the same products are formed from the reaction of DAP with 2'-or 3'-phosphate nucleotides **96**-2'P and **96**-3'P. 5'-nucleotide-amidophosphates **97** are formed from the reaction of DAP and 5'-phosphate nucleotides **96**-5'P.*

c: *The reaction of glycerol **75**, nonanoic acid **99** and DAP provides a cyclophospholipid **98** which is capable of forming liposomes in aqueous solution.*

Methods for phosphorylation of nucleosides and nucleotides (***Figure 2.9b***) and lipid precursors (***Figure 2.9c***) using DAP have recently been published by Krishnamurthy and colleagues.[208] Nucleophilic attack of the 2′- or 3′-OH of nucleosides **96** on protonated DAP followed by cyclisation provides the 2′,3′-cyclic phosphate **96**-2',3'cP. The 2′- or 3′-nucleotide monophosphates (**96**-2'P or **96**-3'P) also react with DAP to give the corresponding 2′,3′-cyclic phosphates **96**-2',3'cP; the 5′-monophosphates **96**-5'P react with DAP to afford the corresponding 5′-nucleotide-amidodiphosphates **97**.[208] A cyclophospholipid **98 –** formed by the reaction of glycerol **75**, nonanoic acid **99** and DAP – is capable of forming stable giant vesicles in aqueous solution.[208]

The ubiquity of DAP as a phosphorylation reagent across several classes of biologically important molecules, and the specificity of its mechanism, demonstrate its importance as a useful potentially prebiotic molecule.

2.1.4 Other methods of prebiotic peptide oligomerisation

Common reagents used in peptide synthesis have been suggested as models for prebiotically plausible reagents: EDC **22**, for example, has been proposed as a model for activating agents acting by attack of the carbonyl group, such as cyanamide.[124,273,288] **22** can be used to oligomerise ß-amino acids (the carbonyl group is attacked by the carbodiimide), but not α-amino acids, as the carbonyl group of the dipeptide can become activated and cyclise to form diketopiperazine **100** (***Figure 2.10***).[273] Conversely, α-amino acids can be activated by carbonyl diimidazole **101**, *via* the formation of an *N*-carboxy anhydride (NCA, **17**) intermediate, in the presence of CO_2: the intermediary **17** is formed when the carbonyl group of the amino acid is activated. ß-Amino acids, however, cannot be activated using this NCA method, unless they contain an α-carboxylate group, as activation would occur *via* the formation of a slower-forming 6-membered NCA.

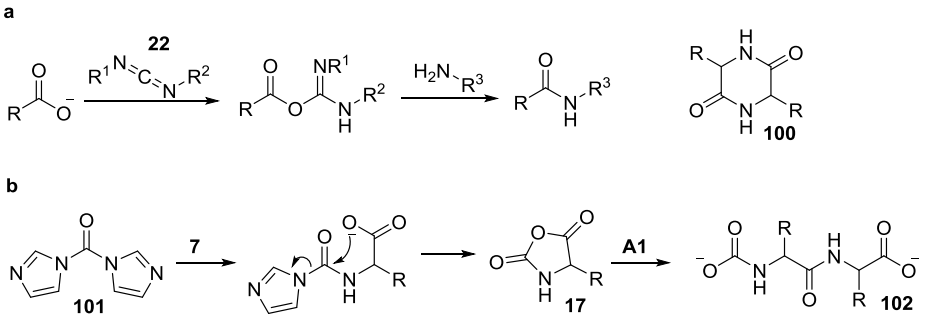


Figure 2.10*: activation of amino acids with EDC* ***22*** *(R^1=Et, R^2=$(CH_2)_3N(CH_3)_2$) and CDI* ***101****.*
a) *Amide bond formation via EDC-activation of the carboxylic acid moiety. Diketopiperazine is a byproduct of the EDC-mediated oligomerisation of α-amino acids*
b) *Activation of an amino acid* ***7*** *with* ***101*** *proceeds via formation of* ***17****. Ring-opening with an amino acid yields an N-acylated dipeptide* ***102****.*

The cyanamide-mediated oligomerisation of glycine has been attempted, but is limited by the tendency of the oligomer to cyclise.[289,290] This can be prevented when an *N*-acyl-α-amino acid (eg **103**) is used instead, with reaction occurring through the formation of a reactive 5(4*H*)-oxazolone cyclic intermediate **104** (***Figure 2.11***).[290]

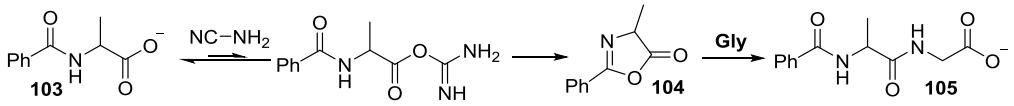


Figure 2.11*: activation of N-benzoyl alanine* ***103*** *with cyanamide results in the formation of cyclic intermediate* ***104****. Addition of glycine* ***Gly*** *yields the N-benzoylated dipeptide.*

17, the active intermediate when **101** is used,[273] is also formed with other, more prebiotically plausible, activating agents. Amino acids **7** can be converted into *N*-carbamoyl amino acids **14** by cyanate; **14** is poorly reactive but can be activated into **17** upon nitrosation (***Figure 2.12***).[291] **17** is also formed as an intermediate in the elimination pathway upon decomposition of **14**.[291]

17 is also formed upon the reaction of amino acids with carbonyl sulfide (COS, ***Figure 2.12***).[73] **17** can then oligomerise – if attacked by a free amine moiety of an amino acid or peptide – or hydrolyse in aqueous solution.[73] Hydrolysis is catalysed by inorganic phosphate (**17** can be formed from the reaction of amino acids and COS in the presence of inorganic phosphate[292]) *via* the transient formation of an aminoacyl phosphate **74**,[293] which indicates that NCAs could also be used as phosphate-activating agents. Aminoacyl phosphates are carboxylic-phosphoric mixed anhydrides that can potentially transfer phosphoryl groups;[293] in fact, the phosphorylating ability of such mixed anhydrides has previously been published.[294] Phosphoryl transfer is central to biology, providing the chemical basis for many fundamental processes of life: phosphorylation is required for the transfer of genetic information, the formation of metabolites, and also the driving force for metabolism through the coupling of chemical energy.[295]

17 has also been shown to aminoacylate nucleotides,[292,296] in a mechanism that bears a resemblance to extant amino acid activation.[274] Biological activation involves the formation of an acyl-5′-phosphate, aminoacyl adenylate **76**, then intermolecular transfer of the amino acid to the 2′- or 3′-terminal adenosine of a tRNA, although the spontaneous intramolecular transfer of the amino acid from the 5′-mixed anhydride to give the 3′-ester has been reported.[297] Reaction of **17**[296] (which can be formed in situ[292]) with adenosine-5′-monophosphate **55**-5'P results in initial aminoacylation giving a phosphoric anhydride **76**, but this is reportedly susceptible to hydrolysis,[292] and desired intramolecular transfer is prevented by competing hydrolysis and polymerisation of **17**.[296] Reaction of **17** with a 3′-nucleotide **96**-3'P with subsequent

intramolecular transfer of the amino acid to 2′-OH has, however, been reported.[296] This could represent the origin of amino acid activation, with the evolutionary path eventually leading to coded protein synthesis. **17** also mediates the formation of fatty acylated amino acids and peptides,[298] thereby linking these key intermediates in the formation of peptides to the chemistry of membranes.

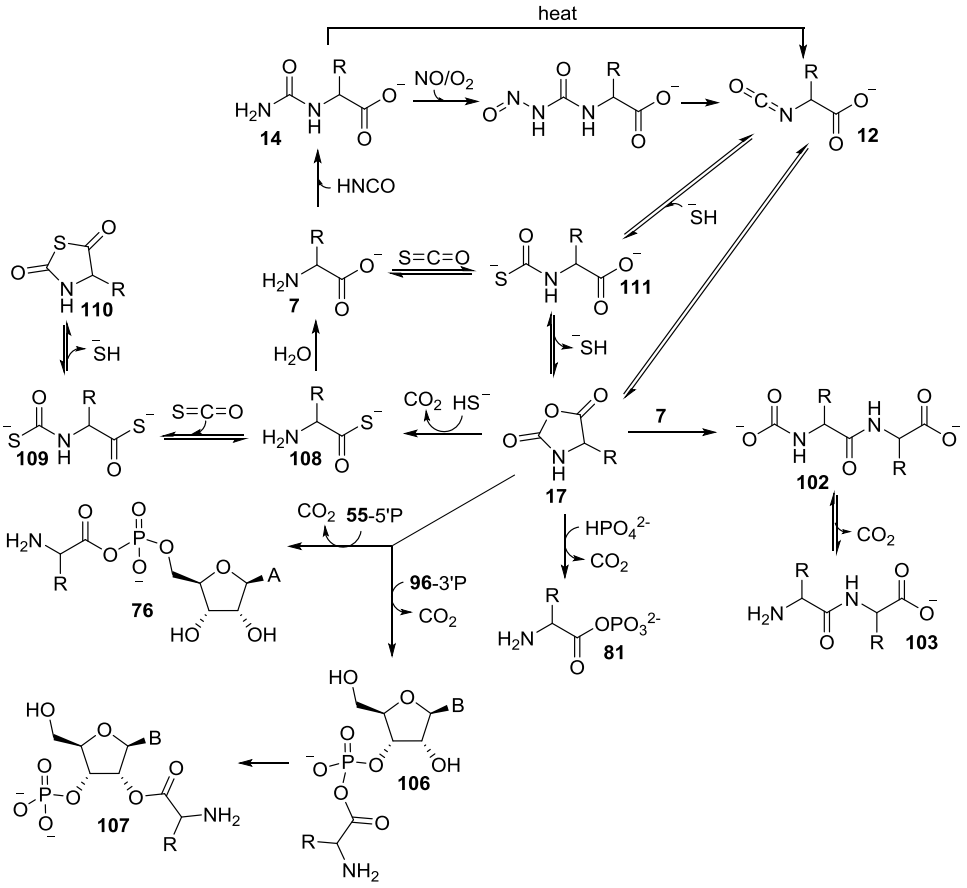


***Figure 2.12**: reaction network of NCA **17**. **17** can be formed from the activation of amino acid **7** with cyanate via the formation of **14** and isocyanate **12**, or from reaction with COS via aminoacyl thiocarbamate **111**. Reaction of **17** with amino acid **7** generates an N-acetylated dipeptide **102** which can be hydrolysed to dipeptide **79**. Reaction of **17** with inorganic phosphate yields aminoacyl-phosphate anhydride **74**; reaction with **55**-5'P generates aminoacyl adenylate **76**, while reaction with **96**-3'P provides mixed anhydride **106** which can undergo intramolecular rearrangement to give ester **107**. Reaction of **17** with H_2S provides α-aminothioacid **108** which can be activated by COS via **109** to give **110** (analogous to **17**).*

The addition of sulfide ions to **17** results in the formation of an α-amino thioacid **108**, albeit in very small (< 2%) yields.[299] Peptides are also observed among the products of the reaction of amino acids with CS_2 – although it is possible that transiently formed COS is in fact the active reagent.[300]

The formation of α-amino thioacids is important, as these have been suggested as opening a route to amino acid oligomerisation.[299,301,302] An amide bond can be formed from oxidative acylation by thioacids,[302] although oligomerisation of an α-amino thioacid (the thioacid derivative of glutamic acid) on a primer proceeds only very slowly.[301] This oligomerisation is believed to occur *via* the formation of **17**, as addition of bicarbonate accelerates the reaction, but cyclisation competes with linear peptide synthesis.[301] Glycine thioesters are also capable of auto-condensation, but, again, cyclisation is a competing process.[303] α-Amino thioesters can be produced from the reaction of an aldehyde with ammonia and a thiol;[304] thioacids could be produced from the reaction of H_2S with nitriles.[302]

An activation of amino acids on early Earth could therefore be avoided if the aminonitrile intermediates could be instead converted into amino acid-derivatives more reactive than the natural amino acids. The carboxylic moiety of the amino acids represents the lowest part of the free energy in the reaction coordinate of highly oxidised carbons. The kinetically stable nitrile moiety of the aminonitrile intermediates could be used as a source of regiospecific chemical activation that could be used in peptide ligation, which is otherwise lost in hydrolysis to amino acids. The direct condensation of aminonitriles is very inefficient, however.[305]

Intramolecular activation of α-amino nitriles has previously been studied by aminal-promoted hydrolysis, which leads to the formation of amino acids,[59,60,270] and by thiolyis.[72,306] The mechanisms of both these reactions proceed *via* a five-membered cyclic intermediate, structurally similar to the method of activation by cyclic trimetaphosphate (***Figure 2.13***). Thiolysis does partially retain the chemical energy lost to hydrolysis, but thioesters and thioacids ultimately cyclise instead of forming linear peptides.[72]

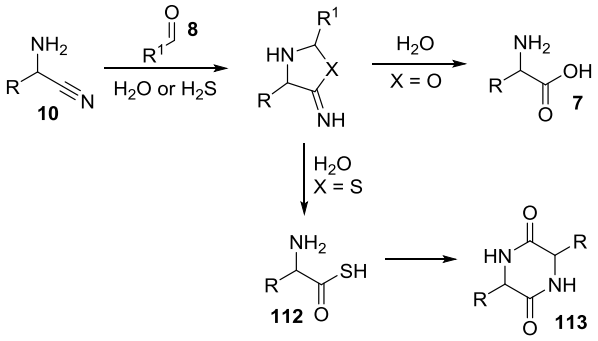


Figure 2.13*: Intramolecular aminonitrile activation. Reaction of* ***10*** *with an aldehyde results in aminal-promoted hydrolysis, forming amino acids. The addition of hydrogen sulfide results in the formation of a five-membered cyclic intermediate which is hydrolysed to give the thioacid* ***112****.*

2.2 Aims

We sought to explore the potential of DAP to direct the intrinsic reactivity of amino acid derivatives in water, ultimately leading to peptide bond formation. This will help us elucidate the predisposed chemical pathway leading to the definition of biological form and function, which chemically couples peptide synthesis to phosphate activation.

It was hypothesised that modification of the Strecker reaction – replacing ammonia with DAP – would lead to the regiospecific tethering of the amidophosphate to a nitrile. The aminonitrile framework would therefore be installed, but with the introduction of a moiety that could activate the nitrile towards nucleophilic attack. Intramolecular trapping of the nitrile would result in the production of an azaphospholidine **115** (***Figure 2.14***): a moiety structurally similar to the activated cyclic phosphoramides elucidated by Rabinowitz, discussed above. The enthalpy stored in the nitrile bond would be used in this activation strategy, rather than being lost during hydrolysis and requiring a further carboxyl-activation step using a condensation agent to recover the lost energy.

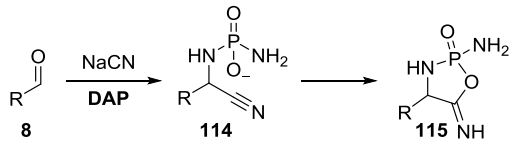


Figure 2.14*: Hypothesised formation of activated peptide bond precursor* ***115.***

It is possible that intramolecular trapping of the nitrile would not occur spontaneously, and that this intermediate would need to be activated in some way to allow the formation of **115**. Initial work focused on investigating the scope of this phosphoro-Strecker reaction.

2.3 Results and discussion

2.3.1 Exploring the optimal reaction conditions

Two model aldehydes – propanal **8d** and isobutyraldehyde **8e** – were selected for studies to explore the optimal conditions of the phosphoro-Strecker reaction, to ensure that any trends observed were not specific to just one substrate. **8d** is not a proteinogenic amino acid precursor, but both **8d** and **8e** are commercial and easy to handle.

2.3.1.1 Exploring the optimal reaction conditions: pH

It is known that, in the presence of excess cyanide, the rate-determining imine formation step of the Strecker reaction depends on the concentration of amine in its neutral form.[307] Although our attempts to determine the pK_a of DAP by ^{31}P NMR spectroscopy were not entirely successful (DAP hydrolysis and condensation at lower pHs meant the chemical shift of DAP could not be accurately determined across the full pH range required), our measurements nevertheless suggested that DAP's pK_a = 5.5: significantly lower than the pK_a of ammonia (9.2). This, together with the known imine-tethered phosphoryl-transfer reactivity exhibited by DAP, suggests that DAP would be available in its neutral form at lower pH values, and so the phosphoro-Strecker reaction could take place in significantly more acidic solution than the conventional Strecker reaction. DAP was also predicted to out-compete ammonia (if present) in Strecker reactions at acidic and neutral pH. On the other hand, it is known that the formation of *N*-substituted aminonitriles is influenced by the steric and inductive effect of the amine substituents, including the number of protons and alkyl groups carried by the amine:[270] therefore the steric bulk of DAP compared to ammonia may reduce its reactivity, even at lower pHs. The impact of reaction pH on the phosphoro-Strecker reaction was duly investigated (***Figure 2.15***). As the Strecker reaction with ammonia must take place under more alkaline conditions, it appeared prudent to explore the extent to which a pH switch between these two reactions could take place, and whether such a switch would also operate when both amines were present in the reaction mixture.

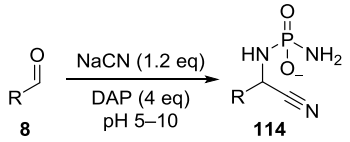
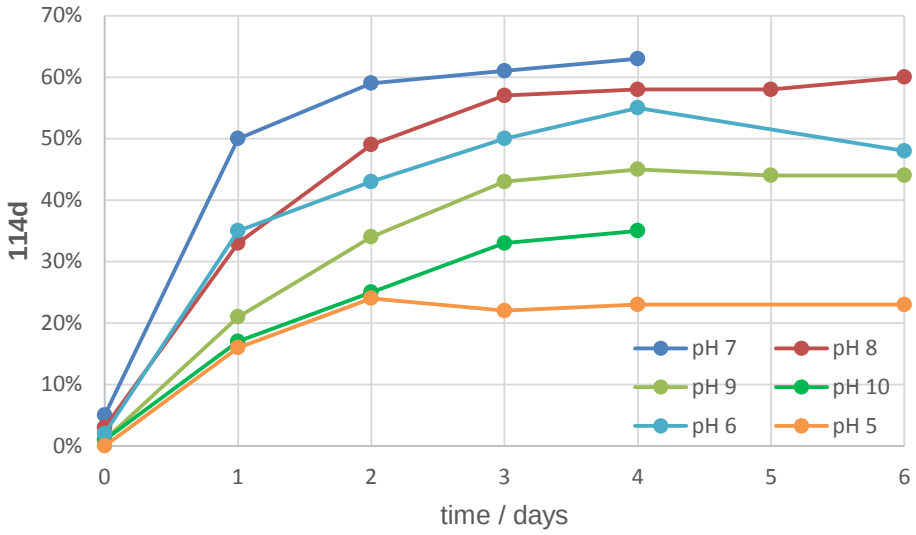


Figure 2.15**: Phosphoro-Strecker reaction of propanal (**8d**, R = CH_2CH_3) and isobutyraldehyde (**8e**, R = $CH(CH_3)_2$), yielding the corresponding product **114.

The optimum pH for the phosphoro-Strecker reaction of both aldehydes was found to be pH 7 (***Graph 2.1*** shows this phenomenon for propanal **8d**): the reaction rate and yield were suppressed at pH values above and below neutral. It is known that ammonia is released during the phosphorylation of α-hydroxy aldehydes, raising the solution pH and stalling reaction progress.[153,154] The pH of the reaction mixture was continuously monitored during investigations of the phosphoro-Strecker reaction and maintained (+/- 0.5 pH units) by addition of HCl.



Graph 2.1*: Reaction of propanal* ***8d*** *(200 mM) with NaCN (1.2 eq) and then DAP (4.0 eq) at a range of pH values. Yields calculated by relative integration of* ^{31}*P NMR spectra: this method means that the yields for reactions at pH 5 and pH 6 are inaccurate after 3 days.*

Quantifying the species present in the reactions at pHs below neutral proved difficult: reactions with both aldehydes at pH 5, and the reaction of propanal at pH 6, started to precipitate after 3 days. NMR analysis of the precipitate revealed that both the internal standard, pentaerythritol, and phosphorylated products were present in all cases. Therefore, calculations of yields based on relative integration to the internal standard in ^{1}H NMR spectra and by relative integration of species seen in ^{31}P NMR spectra at these pH values are inaccurate, but are nevertheless useful in illustrating a general trend in yields over the course of the reaction. The addition of co-solvent was avoided as Strecker reactions are equilibrium-controlled and co-solvents are known to alter the hydration equilibrium of aldehydes in water.[308]

The phosphoro-Strecker reactions were slower and lower-yielding than anticipated, but some reasons for this became apparent on monitoring ^{1}H and ^{31}P NMR spectra. A

range of unidentifiable by-products were observed on the spectral baseline between δ 0.75-1.25 ppm across all pH values (5-10), but particularly prevalent at basic pHs (>8), indicated possible aldolisation of the aldehyde (this complex mixture of (aldol-type) by-products formed in the phosphoro-Strecker reaction of propanal **8d** is highlighted in red in ***Figure 2.16***).

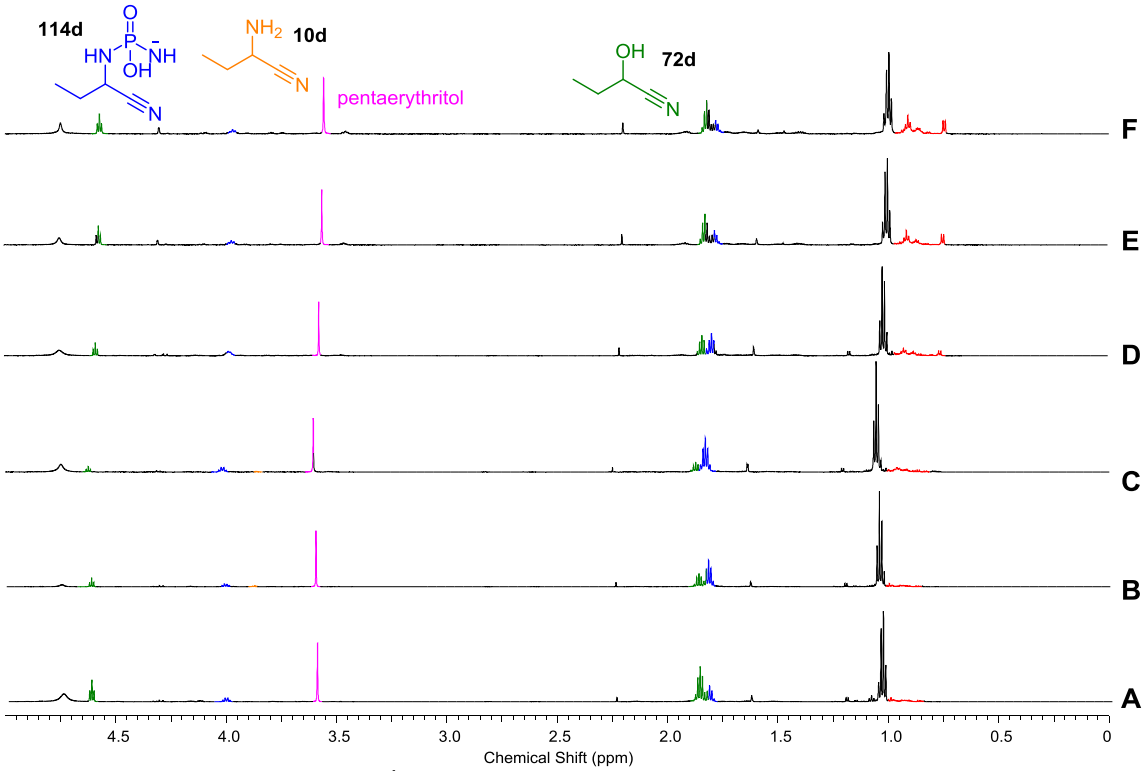


Figure 2.16: *Water-supressed ^{1}H NMR spectra (H_2O:D_2O 9:1, 0.00–5.00 ppm, 700 MHz) showing the reaction of propanal **8d** (200mM) with sodium cyanide (1.2 eq.) and DAP (4 eq.) at various pHs at room temperature after 1 d: **A**) pH 5; **B**) pH 6; **C**) pH 7; **D**) pH 8; **E**) pH 9; and **F**) pH 10. The complex mixture of products formed, especially at higher pHs, indicated by peaks in the methyl region (δ 1.0-0.7 ppm) coloured red.*

Whilst aldolisation is supressed at pH <7,[154] the reaction is nevertheless slow. This is perhaps due to a combination of DAP protonation and acid-catalysed hydrolysis of DAP, reducing its stoichiometry. The small proportion of DAP remaining in the reaction mixture of propanal **8d** at pH 5 can be seen in ***Figure 2.17***; the difference in pH also results in amino protonation and consequent upfield shift of the ^{31}P resonance for DAP.

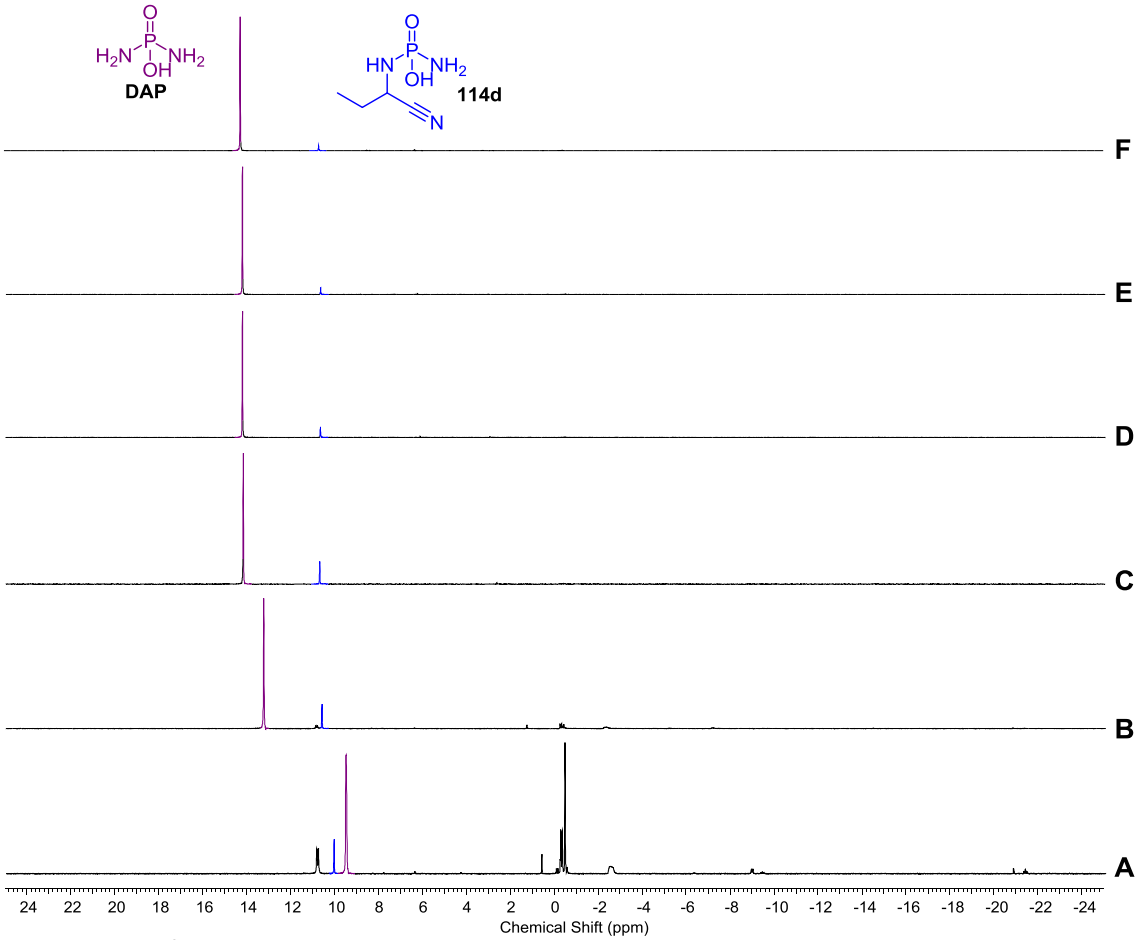


Figure 2.17: *^{31}P NMR spectrum (H_2O:D_2O 9:1, -25.00–25.00 ppm, 284 MHz) showing the reaction of propanal* ***8d*** *(200mM) with sodium cyanide (1.2 eq.) and DAP (4 eq.) at various pHs at room temperature after 1 d:* ***A****) pH 5;* ***B****) pH 6; and* ***C****) pH 7;* ***D****) pH 8;* ***E****) pH 9;* ***F****) pH 10.*

The utility of the pH-dependence of this reaction was tested by performing the equivalent Strecker reactions (using ammonia as the amine, instead of DAP) at pH 7 and pH 10. Very little aminonitrile was observed when using ammonia at pH 7 (6% of **10e** after 3 days in the case of isobutyraldehyde **8e**, and 13% of **10d** after 5 days for propanal **8d**). It was apparent that a complex mixture of by-products formed, as observed in the ^{1}H NMR spectra (shown for the reaction of propanal **8d** in ***Figure 2.18***).

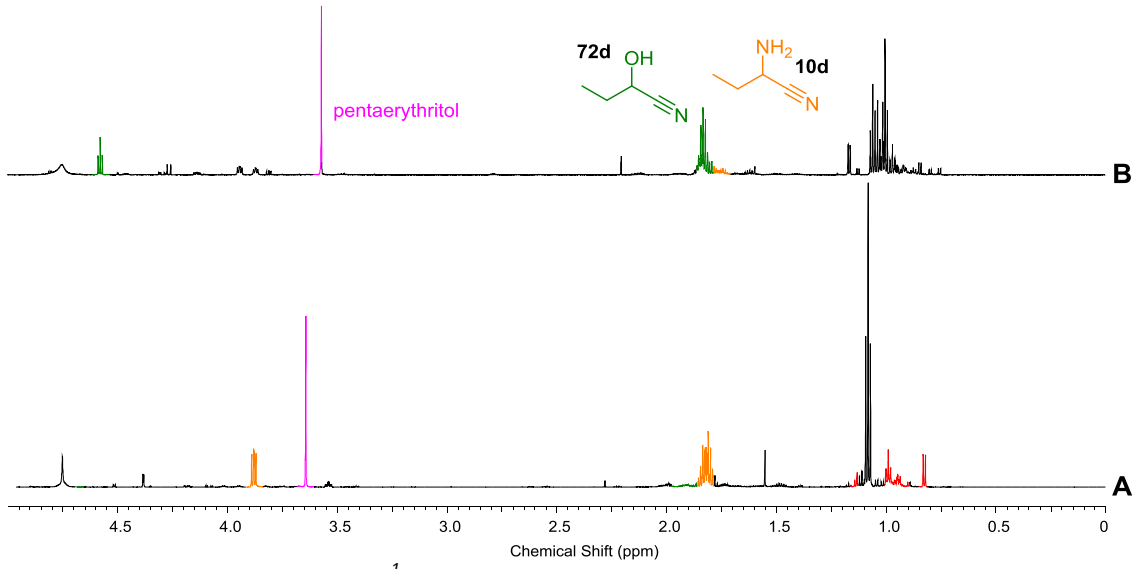


Figure 2.18: *Water-supressed ^{1}H NMR spectra (H_2O:D_2O 9:1, 0.00–5.00 ppm, 700 MHz) comparing the reaction of propanal* ***8d*** *(200mM) with sodium cyanide (1.2 eq.) and NH_4Cl (4 eq.) at room temperature at:* ***A****) pH 10 after 1 d (reaction complete); and* ***B****) pH 7 after 6 d. The complex mixture of products formed can be clearly seen in the methyl region (δ 0.75-1.25) of the reaction at neutral pH; suspected aldolisation products are also present in the reaction at pH 10 (coloured red), albeit to a far less significant extent. The difference in pH has caused the slight difference in chemical shift for the pentaerythritol standard.*

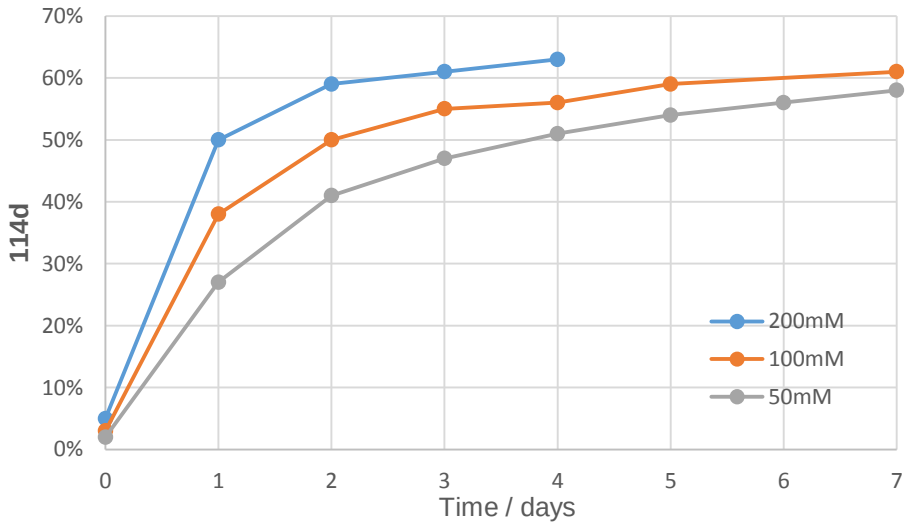
Competition reactions between the ammonia and DAP were also explored to confirm that the phosphoro-Strecker product would be formed at pH 7 even in the presence of ammonia (***Table 2.1***). In competition reactions between DAP and ammonia, the phosphoro-Strecker product **114** dominated at neutral pH, and the unphosphorylated aminonitrile **10** was the major product at pH 10. Disappointingly, given the 4 unit pK_a difference between ammonia and DAP, there was not a complete selectivity switch at any pH investigated: some phosphoro-Strecker product **114** was formed in the presence of ammonia at pH 10, and yields of **114** were supressed in the presence of ammonia at pH 7.

aldehyde	pH	Reaction time / d	Yield **114** + **10**	Ratio **114** : **10**
Propanal **8d**	10	2	50%	1 : 9
	7	6	57%	7 : 1
Isobutyraldehyde **8e**	10	3	46%	1 : 14
	7	8	51%	4 : 1

Table 2.1*: Table showing yields and ratios of phosphoro-Strecker product* ***114*** *and unphosphorylated aminonitrile* ***10*** *when reaction carried out with NaCN (1.2 eq), NH_4Cl (4 eq) and DAP (4 eq) at pH 7 and pH 10.*

2.3.1.2 Exploring the optimal reaction conditions: concentration

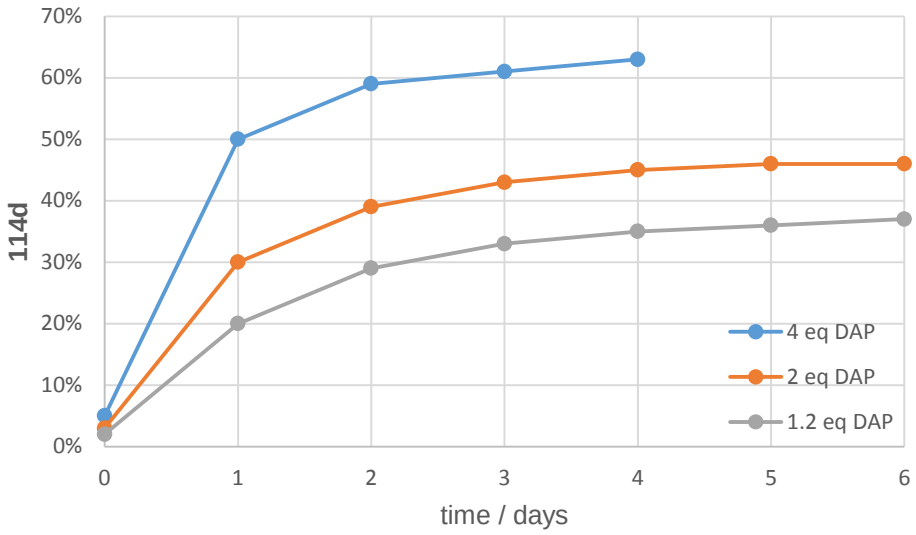
With pH 7 confirmed as the optimal pH for the phosphoro-Strecker reaction, exploration of the effect of other variables was undertaken. Our investigations indicated that higher aldehyde concentrations appeared to be the optimal for the phosphoro-Strecker reaction with 4 eq. of DAP. ***Graph 2.2*** shows the effect of propanal **8d** concentration on the rate of formation of the desired product **114d**: similar yields of **114d** were observed at 50 mM as at 200 mM, albeit after double the reaction time. The case for isobutyraldehyde **8e** was different, demonstrating the pronounced effect of side chain ß-branching on the reaction rate: similar yields could be achieved at 100 mM and 200 mM (albeit after longer reaction times for the former concentration), but the reaction at 50 mM was far slower, not even achieving completion after 14 days.



Graph 2.2*: Reaction of propanal* ***8d*** *(50-200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) at pH 7. Yields calculated by relative integration of* ^{31}*P NMR spectra.*

2.3.1.3 Exploring the optimal reaction conditions: DAP stoichiometry

It has previously been reported that excess ammonia is required for the Strecker reaction to proceed efficiently, avoiding the formation of cyanohydrin, imino-dinitrile and –trinitrile by-products,[309] and that the rate-determining imine formation step of the Strecker reaction depends on the concentration of amine.[307] The effect of DAP stoichiometry on the rate of the phosphoro-Strecker reaction was therefore investigated, indicating that 4 eq of DAP resulted in faster formation of **114** than lower stoichiometires (***Graph 2.3***).



***Graph 2.3**: Reaction of propanal **8d** (200 mM) with NaCN (1.2 eq) and DAP (1.2-4.0 eq) at pH 7. Yields calculated by relative integration of ^{31}P NMR spectra.*

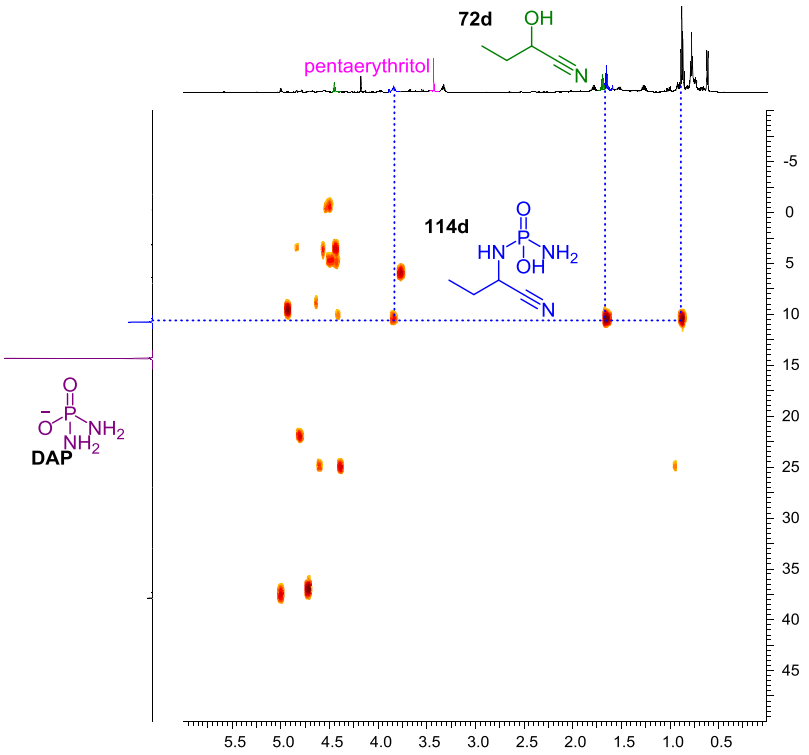


***Figure 2.19:** ^{1}H-^{31}P HMBC NMR spectrum (H_2O:D_2O 9:1) showing the reaction of propanal **8d** (1.60 M) with sodium cyanide (1.2 eq.) and DAP (0.5 eq.) at pH 7.0 after 5 d at room temperature. A number of phosphorylated by-products are present; the high aldehyde concentration has also caused aldolisation, evidenced by the multiple peaks upfield of δ 1.0.*

The requirement for excess DAP for this reaction to proceed mirrors the known excess of ammonia in conventional Strecker reactions.[310] It was thought that perhaps the concentration of DAP – rather than an excess – was a requirement for the progression of the reaction, and therefore this was investigated, using propanal (800 mM) with stoichiometric DAP and NaCN. Unfortunately, this approach proved ineffective: just 16% yield of the desired product was seen after 5 days, and a variety of other phosphorylated and un-phosphorylated products also formed. Attempting to push the reaction further – with an excess of aldehyde – was equally unsuccessful: the mixture of products formed in the reaction of 1.6 M propanal can be clearly seen in the ^{31}P HMBC spectrum shown in ***Figure 2.19***.

2.3.2 Competition between aldehydes and ketones

Natural amino acids all have one unifying structural feature: alongside the α-amine and α-carboxylic acids groups, they universally have a proton attached to the α-carbon. This stands in stark contrast to the amino acids observed in the Miller-Urey reaction or in meteorites, where both α-substituted and α,α-disubstituted amino acids are prevalent.[23,311] The abiotic synthesis of α,α-disubstituted amino acids simply demands a ketone, instead of an aldehyde, in the (phosphoro-)Strecker reaction and is therefore not difficult to envisage prebiotically. It is known that the equilibrium for the formation of cyanohydrin from a ketone and cyanide is not as favourable as the cyanohydrin equilibrium for a (similar molecular weight) aldehyde,[64,312] but α,α-disubstituted aminonitriles have been reported to form in the similar (high) yields as those formed from aldehydes in Strecker reactions.[266]

Competition experiments between ketones and aldehydes were accordingly investigated: 3-methylbutanone **115** was reacted in competition with isobutyraldehyde **8e**, and acetone **74** in competition with propanal **8d** (***Figure 2.20***). In both cases, the aldehyde was observed to out-compete the ketone, although a small amount of the α,α-disubstituted phosphordiamidate aminonitrile (**116** and **117**) still formed.

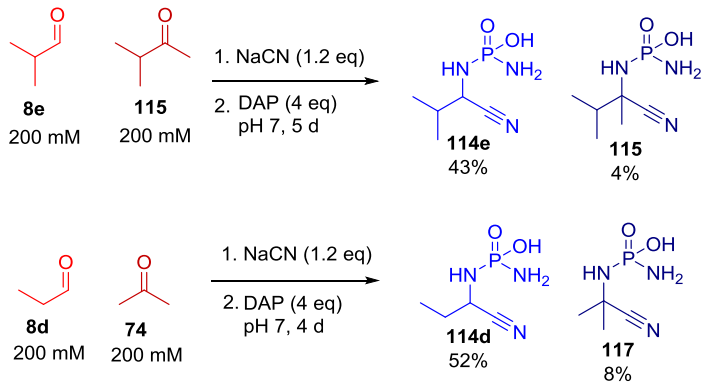


Figure 2.20: *Competition reactions of aldehydes with ketones.*

It is possible that α-H amino acid selection (over disubtituted equivalents) occurred through selective post-Strecker transformations: for example, proteinogenic α-H aminonitriles are more easily hydrolysed to the corresponding amino acids than the α,α-disubstituted aminonitriles.[313,314] [315] A prebiotically plausible method for the selective formation of proteinogenic (α-H) aminonitriles *via* a physiochemical process from a mixture of aldehydes and ketones in the presence of 2-aminothiazole, however, has recently been published.[266] Only the aldehydes are capable of forming a crystalline aminal with 2-aminothiazole, meaning the ketones can be removed with the supernatant (or 'washed away' in a geochemical scenario on the early Earth). Resuspension of the aminals in water, followed by the addition of cyanide and ammonia results in the formation of aminonitriles.[266] The chemical selection of the phosphoro-Strecker reaction offers an alternative or complementary route to synthesis of proteinogenic aminonitriles without the need for aminal-formation and ketone-removal.

2.3.3 Reaction scope

With the optimal conditions for the reaction established, the scope of the reaction was explored using a variety of aldehydes, forming both proteinogenic and non-proteinogenic derivatives. These results confirmed that the introduction of DAP facilitates a highly selective Strecker reaction of aldehydes to produce *N*-phosphordiamidate aminonitriles in moderate to good yields.

Pleasingly, the proteinogenic amino acid precursors generated the corresponding *N*-phosphordiamidate aminonitriles in generally good yields (***Table 2.2***, entries 2-4, 9; 55-76%). Interestingly, the non-proteinogenic derivatives explored were generally formed

in lower yields (for example, entries 5-8): sterics and solubility are likely contributing factors to this yield disparity. It should be noted that **114d** was formed in good yield from the phosphoro-Strecker reaction of propanal **8d**: a short chain, water soluble non-proteinogenic aldehyde.

Entry	Aldehyde	product	Yield (by relative integration ^{31}P NMR spectra)	Time / d
1	Propanal **8d**	**114d**	63%	4
2	Isobutyraldehyde **8e**	**114e**	55%	4
3	2-Methylbutyraldehyde **8f**	**114f**	76%	6
4	3-Methylbutyraldehyde **8g**	**114g**	74%	3
5	Butyraldehyde **8h**	**114h**	37%	4
6	Pentanaldehyde **8i**	**114i**	53%	4
7	2-Ethylbutyraldehyde **8j**	**114j**	35%	8
8[a]	Octanal **8k**	**114k**	33%	7
9[b]	Lactonitrile **72b**	**114b**	74%	3

Table 2.2: *summarising yields (by relative integration of ^{31}P NMR spectra) obtained from the phosphoro-Strecker reaction of various aldehydes (200 mM), in the presence of NaCN (1.2 eq) and DAP (4 eq) in water at pH 7, with periodic addition of HCl (1 M) to maintain the pH.*

[a] Octanal (100 mM), NaCN (1.2 eq) and DAP (4 eq) in 4:1 H_2O:MeCN, pH 7 at start of reaction but thereafter unadjusted.

*[b] The cyanohydrin lactonitrile **72b** was used instead of the corresponding aldehyde acetaldehyde **71.** Lactonitrile (200 mM), NaCN (0.2 eq) and DAP (4 eq) in water at pH 7, with periodic addition of HCl (1 M) to maintain pH 7 +/- 0.5.*

Attempts to form alanine *N*-phosphordiamidate aminonitrile **114b** (entry 9) from acetaldehyde **71** were plagued by the volatility of this aldehyde (boiling point of 21 °C). The phosphoro-Strecker synthesis of **114b** was therefore optimised by using lactonitrile **72b** (the cyanohydrin of acetaldehyde).

Recent reports on the prebiotic synthesis of fatty acetylated amino acids and peptides such as **118** (***Figure 2.21***) inspired us to investigate the phosphoro-Strecker reaction of octanal.[298] *N*-Carboxyanhydride amino acid derivatives **17e** can mediate the condensation of unactivated amino acids and peptides with fatty acids to generate *N*-acylated products (the reaction mechanism is believed to occur *via* nucleophilic addition of the fatty acid carboxylate to the carbonyl of the *N*-carboxyanhydride, generating a mixed anhydride intermediate which is capable of serving as a prebiotic fatty acyl transfer agent).[298] These acylated products can then be incorporated into a proto-cell membrane; it is believed that protocell membranes would have assembled from prebiotically available amphiphiles.[298]

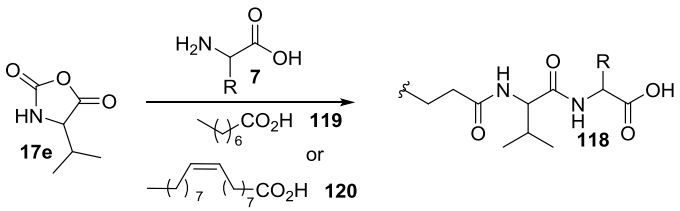


Figure 2.21: *Reaction of valine N-carboxyanhydride* ***17e*** *with amino acids and octanoic acid* ***119*** *or oleic acid* ***120*** *leads to the formation of acylated amino acids and peptides* ***118.***

Alternative amino acid-derived amphiphiles to *N*-acylated amino acids **118** are lipidic α-amino acids with long alkyl side chains. It is known that the zwitterionic nature and hydrogen bonding ability of such amino acids stabilises the formation of large amphiphile aggregates, and the insolubility of the long hydrocarbon side chain is mitigated by the presence of a polar carboxylic acid moiety at a monomeric level.[316]

With this in mind, the phosphoro-Strecker reaction of octanal **8k** was attempted, with 33% of the corresponding *N*-phosphordiamidate aminonitrile **114k** formed after 7 days. The formation of lipidic *N*-phosphordiamidate aminonitriles and their corresponding *N*-phosphonoamino acids, albeit in low yield, could have been important on the early Earth.

Compartmentalisation and localisation of RNA would have been crucial in the RNA world to increase the local concentration of otherwise dilute reactants and prevent the diffusion of reaction intermediates.[317] It has previously been shown that phase separation can result in the compartmentalisation of functional RNAs, facilitating RNA-substrate binding and catalysis, and increasing the rate of reaction.[317] *N*-oleoylarginine (produced from the reaction of arginine, valine- *N*-carboxyanhydride and oleic acid) is spontaneously incorporated into vesicle membranes and increases the local concentration of an RNA oligonucleotide (arginine-rich peptides are known to interact with RNA[318]).[298] Therefore, protocell membranes derivatised with amino acids and peptides could aid RNA catalysis by localising functional RNAs. It is also possible that localisation of catalytic amino acids and peptides at the primordial membrane could facilitate prebiotically important reactions using the same principles; indeed, peptides have been postulated as more efficient catalysts of reactions between membrane associated reagents, than ribozymes made up of anionic and therefore lipophobic RNA.[50] Peptides also have the added advantage of being able to catalyse reactions within the hydrophobic environment of the lipid bilayer interior.[50] Association of a

hydrophobic dipeptide (itself formed from the catalytic action of another dipeptide) with vesicle membranes has been shown to enhance vesicle growth by recruiting fatty acids from surrounding vesicles, thus providing a competitive advantage to the vesicle containing such peptides.[50]

The phosphoro-Strecker reaction of acrolein **121** was also attempted. The production of methionine **Met** and glutamic acid **Glu** in the products of Miller's spark discharge experiment led to Miller and Trump proposing that acrolein could have been an intermediate in the synthesis of these amino acids, and that a variety of other amino acids could have been produced from the reaction of acrolein with other nucleophiles produced *in situ*.[311,319] Michael addition of cyanide to acrolein, followed by the addition of ammonia to the cyanohydrin (2-hydroxypentane dinitrile **123**) generated, results in the Strecker synthesis of an aminonitrile (2-aminopentanedinitrile **124**) which is a precursor to **Glu** and glutamine **Gln** (***Figure 2.22***).[320] It seemed likely that acrolein could be a suitable substrate for the phosphoro-Strecker reaction. Unfortunately, a complex mixture of products were formed, and the presence of the desired *N*-phosphordiamidate aminonitrile in this mixture could not be confirmed.

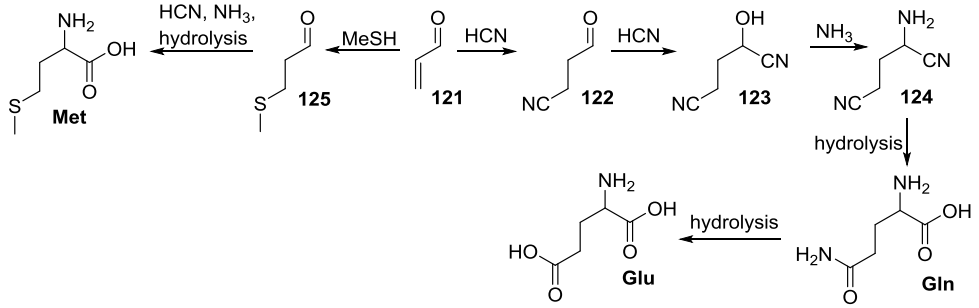


Figure 2.22: *Proposed synthesis of amino acids methionine* ***Met****, glutamic acid* ***Glu*** *and glutamine* ***Gln*** *from acrolein* ***121****. Addition of nucleophiles methanethiol or cyanide to* ***121*** *results in the formation of the Michael addition products* ***125*** *and* ***122*** *respectively. Strecker reaction and hydrolysis results in the formation of the amino acids.*

It is possible that the phosphoro-Strecker reaction could take place on a substrate other than acrolein **121** to afford an *N*-phosphordiamidate aminonitrile that could be derivatised to provide **Glu** and **Gln**. It has been shown that the cyanohydrin intermediate, 2-hydroxypentane dinitrile **123**, can be formed by cyano-sulfidic photoredox elaboration of cyanoacetylene.[64] The phosphoro-Strecker reaction of 2-hydroxypentane dinitrile **123** formed in this way may be more selective than from

acrolein **121**, offering an alternative route to further proteinogenic amino acid precursors.

The phosphoro-Strecker reaction of glycolaldehyde **32** was also explored. Unfortunately, it was immediately clear that the reaction was not as clean as that for the other aldehydes attempted (with the exception of acrolein): ^{1}H-^{31}P HMBC spectra demonstrated that a variety of other phosphorylated species were formed alongside **114c** (***Figure 2.23***). This is not a surprising result: DAP is a known phosphorylating agent for α-hydroxy aldehydes such as glycolaldehyde **32**.[153,154] Additionally, aldolisation by-products have been reported across all pH values (evocatively described by Krishnamurthy as a 'forest of signals between 3.5 and 4.6 ppm' in the ^{1}H NMR spectrum[152]) in the reaction of DAP and glycolaldehyde **32**, although sugar α-phosphorylation by DAP is also known to be pH dependent and is supressed above pH 7. Reaction conditions for the phosphorylation of α-hydroxy aldehydes with DAP have been optimised to increase the rate of formation of the desired products, as well as their yield, whilst reducing the formation of by-products.[154] It was therefore thought that reaction conditions that favoured the formation of the desired phosphoro-aminonitrile **114c** could be found.

A lower concentration was successful in reducing the rate of aldolisation in the phosphorylation of glycolaldehyde **32** with DAP.[154] Reducing the pH of the reaction mixture also increased the rate of formation of glycolaldehyde-2-phosphate (the reaction was complete after just 4 h at pH 4).[154] It was clear that such a low pH would not be successful in the phosphoro-Strecker reaction: protonation of sodium cyanide would make it a less effective nucleophile and result in its volatilisation; moreover, DAP is acid-labile (this phenomenon is mitigated by the fast reaction time of the aldehyde phosphorylation reaction) and so too is the phosphoro-aminonitrile product **114**. The phosphoro-Strecker reaction was therefore attempted at lower concentration in an attempt to reduce the formation of by-products. Although the same pH-dependence was observed (***Graph 2.4***), much lower yields were achieved at 50 mM than 200 mM: the highest yield of **114c** observed at 50 mM was 12% after 19 days, initial pH 7.0, in contrast to 37% after 6 days, initial pH 8.9. It was clear from observation of NMR spectra that a variety of by-products continued to form at 50 mM.

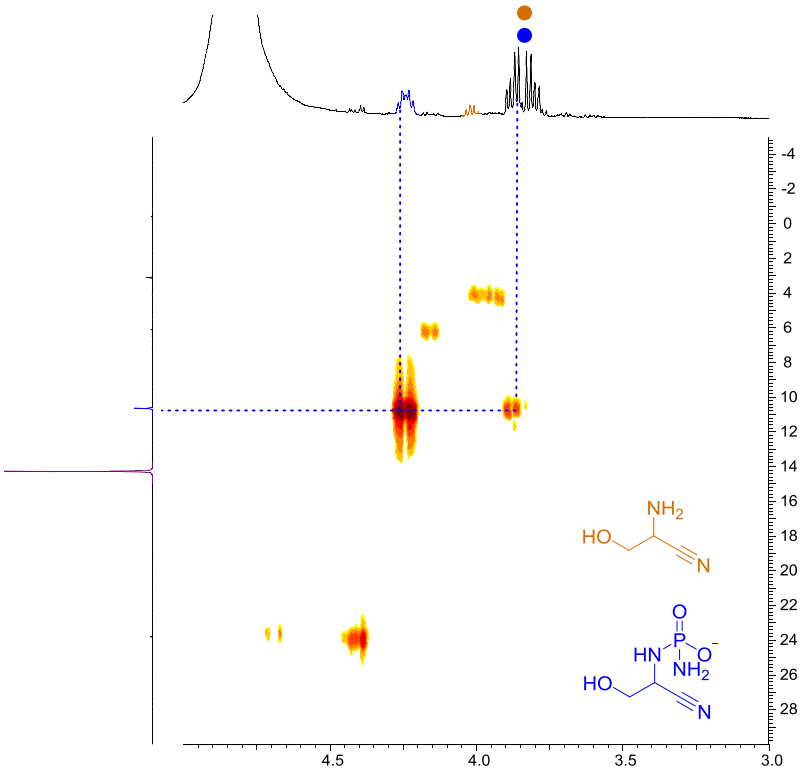
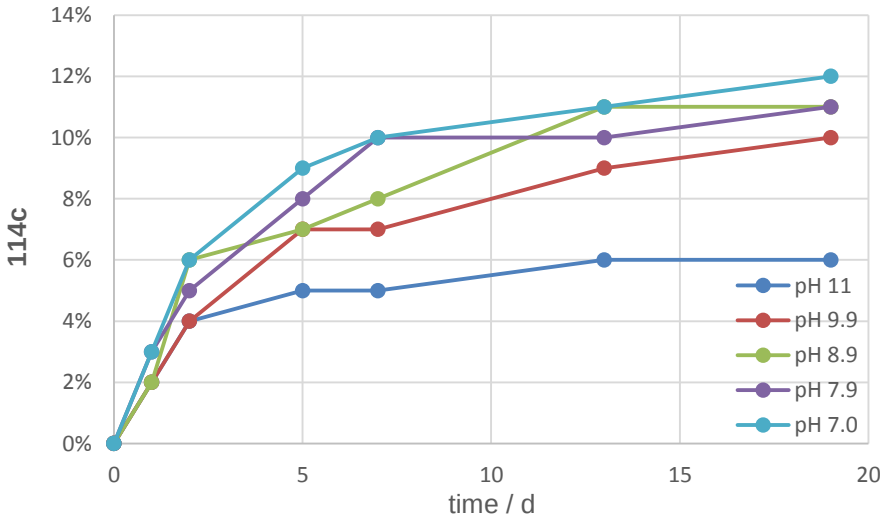


***Figure 2.23**: ^{1}H-^{31}P HMBC showing the reaction of glycolaldehyde **32** (200 mM) with NaCN and DAP (4 eq) after 1 day at room temperature, initial pH 8.9. A variety of phosphorylated by-products are formed alongside the desired product **114c** and unphosphorylated aminonitrile **10c**.*



***Graph 2.4**: Reaction of glycolaldehyde **32** (50 mM) with NaCN (1.2 eq) and DAP (4 eq) at pH 7-11. Yields calculated by relative integration of ^{31}P NMR spectra.*

The low yields shown in this competition reaction drove us to investigate the reaction of DAP with pre-formed aminonitrile **10c**: essentially no reaction was observed, with just 1% yield of the phosphoro-aminonitrile seen after 2 days with 4 eq DAP.

Whilst monitoring phosphoro-Strecker reactions through ^{1}H-^{31}P NMR studies, it became apparent that a second phosphorylated aminonitrile was being formed, albeit in low yields, from the phosphoro-Strecker reaction with some aldehydes (eg 3-methylbutyraldehyde **8g**, 8% after 4 days; pentanaldehyde **8i**, 1% after 4 days). NMR analysis of the reaction mixture suggested this product was structurally very similar to the *N*-phosphordiamidate aminonitrile **114**, but with a significantly more shielded phosphorus atom (δ_P 7.15 in the case of the 3-methylaldehyde derivative, compared to δ_P 10.47 for **114g**). This distinctive chemical shift allowed us to assign it as the *N*-phosphono aminonitrile **126**, produced either from the hydrolysis of *N*-phosphordiamidate aminonitrile **114**, or from the participation of monoamidophosphate (MAP, from the *in situ* hydrolysis of DAP or *N*-phosphordiamidate aminonitrile) in the phosphoro-Strecker reaction (***Figure 2.24***).

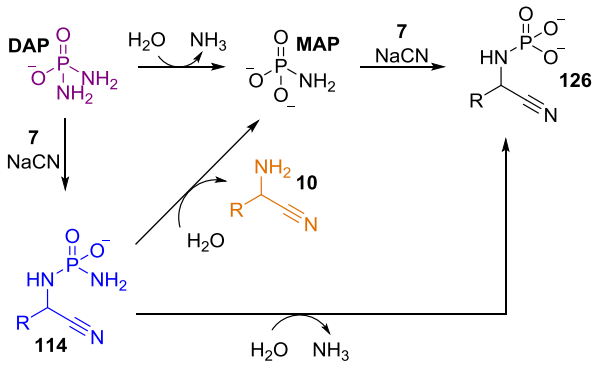


Figure 2.24*: Possible routes to N-phosphonoaminonitrile* ***126*** *from the reaction of aldehyde* ***7*** *with NaCN and DAP.*

2.3.4 Reaction mechanism

It is known that competition between cyanide and the amine in Strecker reactions first favours cyanohydrin formation.[307] It has also previously been observed that addition of cyanide to aldehyde followed by amine results in better yields than an order of addition that reflects the reaction mechanism (ie amine to aldehyde followed by cyanide): indeed, this led to a suggestion that the reaction proceeded *via* a cyanohydrin intermediate.[310] This reported observation has been corroborated by our own observations on the effect of order of reagent addition in the phosphoro-Strecker reaction of 2-methylbutyraldehyde **8f** (5% yield of **114f** when DAP was added to **8f**

before cyanide; 83% when cyanohydrin formed first). The C-O bond in cyanohydrin is strengthened by the presence of a nitrile group and cannot be broken by nucleophilic attack of an amine; a hemiaminal contains a labile hydroxyl group which could in principle be displaced by cyanide (or cyanide could simply attack an imine intermediate).[310]

We postulate that the formation of cyanohydrin **72** is especially important in the neutral pH phosphoro-Strecker reaction, as it exists in the reaction mixture as a cyanide- and aldehyde-sink (***Figure 2.25***). It is known that the dissociation of cyanohydrin is rapid and reversible.[310] Dissociation of cyanohydrin **72** releases aldehyde (which can react with DAP, giving a phosphoro-imine) and cyanide (which can attack the phosphoro-imine **A47**); volatilisation of cyanide and aldehyde under the neutral pH of the reaction is thus minimised.

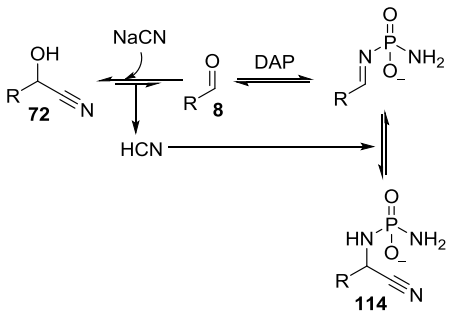


Figure 2.25: *Potential reaction mechanism for the phosphoro-Strecker reaction.*

2.3.5 Purification of *N*-phosphordiamidate aminonitriles

Attempts to purify the *N*-phosphordiamidate aminonitriles formed by ion exchange and by reverse phase flash column chromatography were unproductive. Adaption of a protocol for the purification of *N*-phosphonoamino acids **75**, formed from the reaction of sodium trimetaphosphate and amino acids, by precipitation of the unwanted inorganic phosphates, however, proved successful.[279]

2.3.6 Derivatisation of *N*-phosphordiamidate aminonitriles

Aminonitriles are known to undergo two competing processes in aqueous solution: decomposition into the aldehyde from the sequential elimination of cyanide and ammonia (the retro-Strecker reaction), and hydration of the nitrile to give the corresponding amide **127**, followed by hydrolysis to the amino acid (***Figure 2.26***).[313] Aminonitriles containing an α-H tend to undergo faster hydration,[313] while α,α-disubstituted aminonitriles instead display a propensity to undergo the retro-Strecker

reaction (to the corresponding ketone),[313] particularly in mild acidic media.[314] This difference in reactivity can be used to justify the absence of α,α-disubstituted amino acids from nature: the difficulty of hydrolysis of α,α-disubstituted aminonitriles to the corresponding amino acids meant that only the amino acids containing an α-H were available on the early Earth.[315] We anticipated that the *N*-phosphordiamidate aminonitriles – like unphosphorylated α-aminonitriles – would tend to undergo hydration, and not decomposition. Furthermore, it was hypothesised that the presence of an α-*N*-phosphoryl group could promote hydration, mimicking the aminal-promoted intramolecular hydration mechanism already discussed in the introduction and the introduction to this chapter.

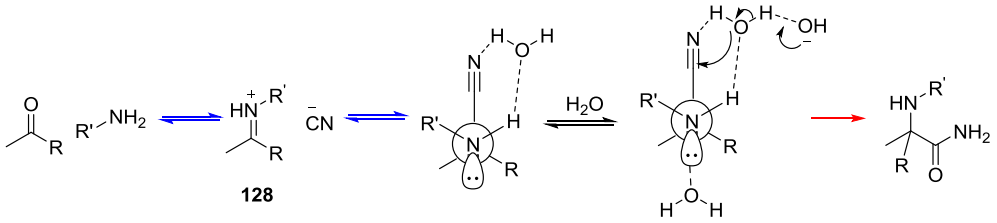


Figure 2.26: *Commeyras and colleagues postulate that trapping a water molecule between the amine proton and the nitrile should promote polarisation of the nitrile bond, facilitating attack by hydroxyl (red arrows). Departure of cyanide (blue arrows) occurs with anchimeric assistance from the nitrogen free lone pair, forming a pi-bonded cationic intermediate* ***128****. When R=H, departure of cyanide must occur from a 2° carbon. The identity of the amine substituents also affects the relative rates of each reaction.*

2.3.6.1 Hydrolysis in acidic solution

The knowledge that exposure of *N*-phosphonoamino acids **75** to low pH liberates the free amino acids and inorganic phosphate,[279] and that hydrolysis of amino nitriles at low pH effects their conversion to amino acids[309,321] appeared to present a method for the synthesis of the free amino acids from the *N*-phosphordiamidate aminonitriles (***Figure 2.27***).

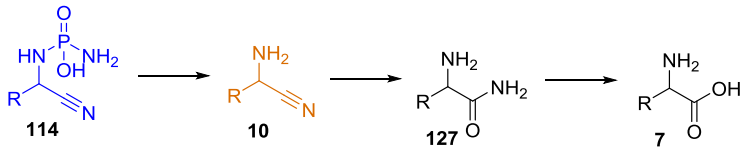


Figure 2.27: *expected reaction of N-phosphordiamidate aminonitriles in acidic solution.*

Hydrolysis at pH 1.5 and 50 °C, however, resulted only in the cleavage of the α-N-P bond liberating the aminonitrile. Treatment of alanine *N*-phosphordiamidate aminonitrile **114b** in this manner resulted in the expected complete hydrolysis of the

phosphordiamidate group after 4 h (***Figure 2.28***); the reaction of isoleucine *N*-phosphordiamidate aminonitrile **114f** was slower, with 10% of the starting material still remaining after 24 h. The acidic hydrolysis of both these derivatives appeared to progress cleanly, with only the *N*-phosphordiamidate aminonitrile and unphosphorylated aminonitrile observed by ^{1}H NMR spectroscopy.

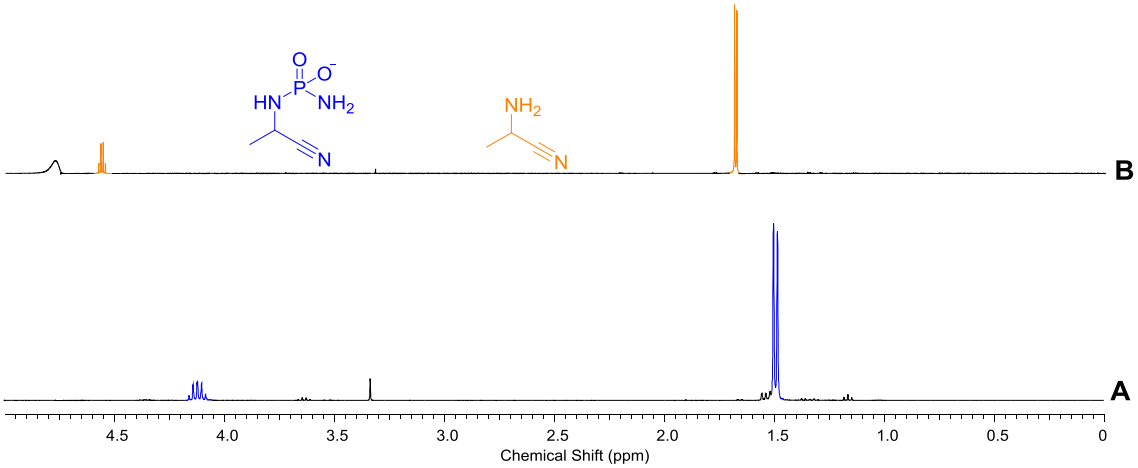


Figure 2.28: *Water-supressed ^{1}H NMR spectra (H_2O:D_2O 1:9, 0.00–5.00 ppm, 700 MHz) to show the reaction of alanine phosphordiamidate aminonitrile* ***114b*** *(100mM) at pH 1.5 at 50°C:* ***A****)* ***114b*** *at pH 6.5* ***B****) after heating at 50°C at pH 1.5 for 4 h.*

Another unphosphorylated product was formed in the reaction of valine *N*-phosphordiamidate aminonitrile **114e**. Our initial assumption was that this was the amino acid or amino amide (and that hydrolysis of the nitrile group was taking place); however, analysis by ^{1}H-^{13}C HMBC (***Figure 2.29***) revealed that both products contained nitrile groups, and there was no cross correlation peak to indicate that hydrolysis of the nitrile to give a carboxylic acid or amide had taken place. Spiking with valine cyanohydrin **72e** confirmed the presence of this species.

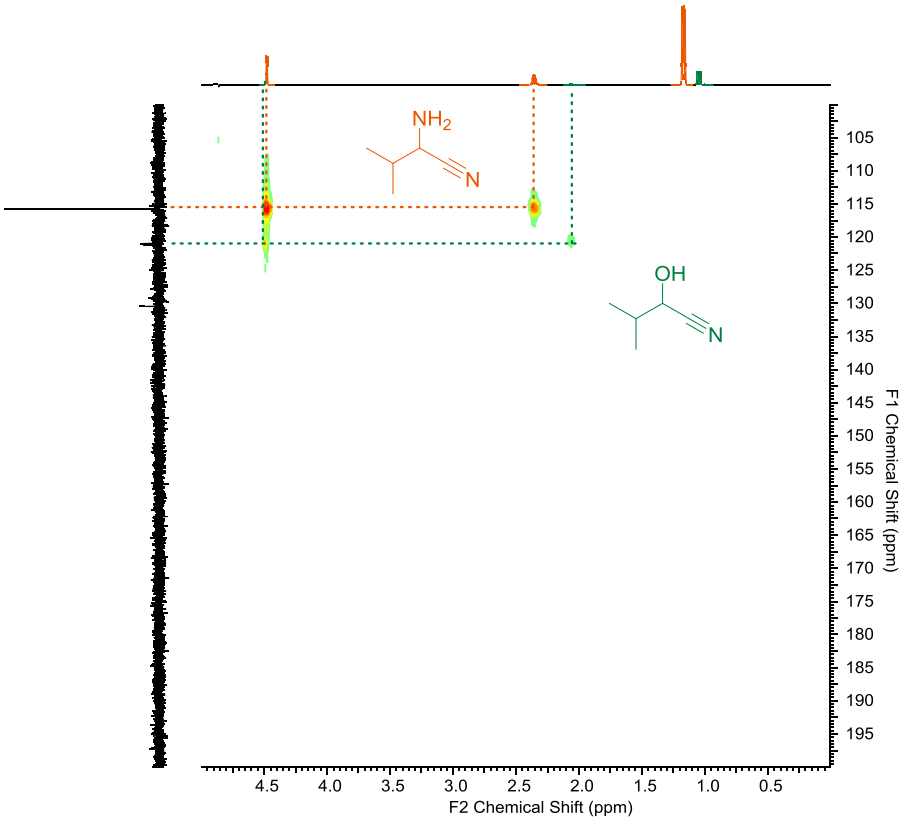


Figure 2.29: *^{13}C-^{1}H NMR HMBC NMR spectrum to show the reaction of valine phosphordiamidate aminonitrile* ***114e*** *(100mM) at pH 1.5 after 22 h at 50 °C. The two products (****10e*** *and* ***72e****) both contain nitrile groups; there is no coupling to indicate that hydrolysis to an amino acid or amino amide (expected in the region δ 170-190) has taken place.*

A 'one-pot' acid hydrolysis at 50 °C was also attempted for each of these aminonitrile derivatives, using unpurified material. On each occasion, the corresponding cyanohydrin was formed alongside the aminonitrile. The one-pot acid hydrolysis of alanine *N*-phosphordiamidate aminonitrile **114b** at room temperature, however, did not result in the formation of cyanohydrin.

The most likely path for the formation of the cyanohydrin appears to be a retro-Strecker reaction of the aminonitrile, followed by cyanohydrin formation from the aldehyde and cyanide released. This reaction mechanism would also account for the sub-quantitative yields calculated by relative integration to an internal standard (for example, just 64% of material could be accounted for in the reaction of isoleucine *N*-phosphordiamidate aminonitrile **114f**, 75% with the valine derivative **114e**). Cyanide produced in a retro-Strecker reaction would certainly volatilise under the low pH and heated reaction conditions, and it is possible that the aldehydes produced would also volatilise. The

reaction mechanism appears to be driven by the presence of an *N*-phosphordiamidate or -phosphoryl group and heat. Control reactions – heating the unphosphorylated aminonitriles at pH 1, and in the presence of phosphoric acid and DAP – did not undergo any change detectable by ^{1}H NMR.

The optimal pH for the conversion of aminonitriles into the corresponding amino acids has been subject to debate in the past. Cocker and Lapworth suggested that the majority of α-aminonitriles revert to cyanide, ammonia or amine (and, presumably, the corresponding aldehyde and ketone, although this is not stated) in the presence of alkali, and therefore that hydrolysis should take place under acidic conditions (specifically, 40% H_2SO_4) to avoid this decomposition,[309] a finding later corroborated by Cook and Cox.[321] Pascal and co-workers, however, noted that the search for mild and selective conditions for the hydration and hydrolysis of aminonitriles led to their studying the behaviour of aminonitriles in weakly alkaline solution. The aminonitriles did not undergo appreciable decomposition, but instead underwent hydration.[59]

2.3.6.2 Hydrolysis in alkaline solution

Our finding that *N*-phosphordiamidate aminonitriles appear to undergo α-N-P bond cleavage but also decomposition in preference to nitrile hydration meant that investigating their behaviour under alkaline conditions was the next obvious step. Hydration and hydrolysis of the nitrile moiety at high pH would also mean the *N*-phosphoryl moiety would not be cleaved.[279]

Alanine *N*-phosphordiamidate aminonitrile **114b** at room temperature was completely consumed after just one day. A species exhibiting a doublet in the ^{31}P NMR spectrum (shown in ***Figure 2.30***) at δ_P 11.96 (J = 10.5 Hz) rapidly formed and was also completely consumed after 1 day. The chemical shift indicates that this phosphorous is coupled to two nitrogen atoms, indicating the *N*-phosphordiamidate moiety is intact. Another species, also an *N*-phosphordiamidate with a doublet at δ_P 12.62 (J = 9.5 Hz) was quickly formed but slowly started decreasing after 1 day (although was still present in the reaction mixture after 33 days). ^{13}C NMR analysis revealed that this species had a deshielded carbon at δ_C 184.0. This is in the region expected for an amino acid cabonyl, although it could also be a very deshielded amide (amides are generally found in the range δ_C 170-180: their carbonyl carbons are more upfield shifted than corresponding carboxylic acids due to the resonance contribution of the nitrogen lone

pair[322]). It seems more likely, however, that the first intermediate formed is the amide derivative **129b**, which is then hydrolysed to the corresponding *N*-phosphordiamidate amino acid **89b** as the second intermediate. Unfortunately, the speed of decomposition of **129b** meant a 2D ^{13}C NMR could not be obtained to confirm this hypothesis. The final product was characterised by a ^{31}P peak (δ_P 8.68, J = 9.5) very close to monoamidophosphate (MAP, observed at δ_P 8.55): This indicates that the final product contains a phosphoryl moiety (instead of a phosphordiamidate) coupled to the amino acid through the α-amine. The carbonyl carbon for this derivative is slightly more deshielded, at δ_C 184.8. On this evidence, the final product was assigned as alanine *N*-phosphonoamino acid, **75b**.

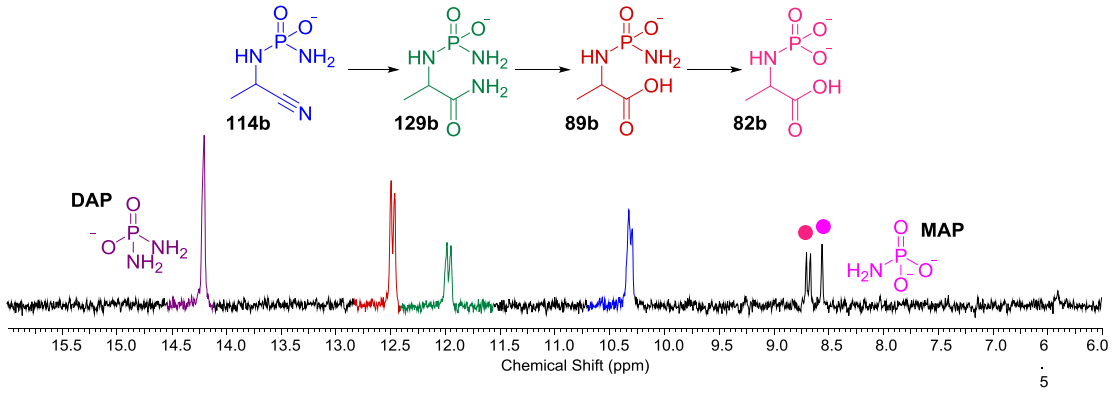
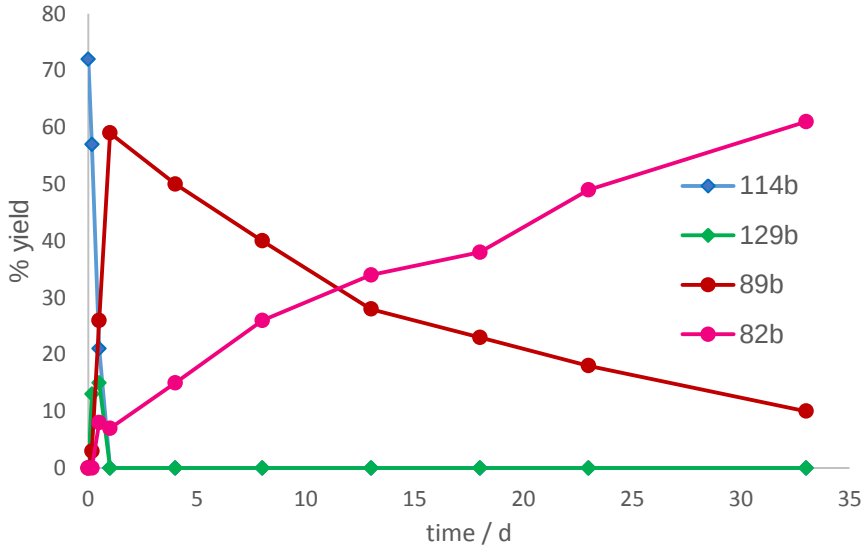


Figure 2.30: *^{1}H-^{31}P coupled NMR spectrum (H_2O:D_2O 1:9, 6.00–16.00 ppm, 284 MHz) showing the reaction of alanine phosphordiamidate aminonitrile (100mM) at pH 13 at room temperature after 12 h. The starting material, final product, and two intermediate species can all be seen.*



Graph 2.5: *Basic hydrolysis of purified alanine N-phosphordiamidate aminonitrile **114b** (100 mM) at room temperature. Yields calculated by relative integration of identifiable species seen by ^{31}P NMR.*

There was in increase in the proportion of DAP (present as a minor contaminant in purified **114b**) and MAP (from the hydrolysis of DAP or the aminonitrile/amino acid derivatives) observed in the reaction over time according to relative integration of all species seen by ^{31}P NMR (***Graph 2.5***). This implies that hydrolysis of the α-N-P bond is occurring, releasing unphosphorylated aminonitrile and amino acid, but none were detected, or that aldehyde was released through a retro-Strecker reaction. It is possible that such derivatives were present in concentrations below the detection threshold of the NMR spectrometer.

When heated at 50 °C, the basic hydrolysis of **114b** proceeded much faster – indeed, the first intermediate (**129b**) could not detected – with complete conversion of **89b** to the *N*-phosphoryl amino acid **75b** after just 4 days. The proportion of DAP and MAP generated over the course of the reaction was more significant than that at room temperature (32% and 10%, compared to 28% and 10% at room temperature): it appears this increased temperature promotes phosphoryl-cleavage. Again, however, unphosphorylated products could not be detected. Pleasingly, the reaction was equally effective on unpurified **114b**, with complete conversion to **75b** observed after just 3 days when heated at 50 °C.

Hydrolysis of valine and isoleucine *N*-phosphordiamidate aminonitrile (**114e** and **114f**) at pH 13 also resulted in the formation of the corresponding *N*-phosphoryl amino acids, but the formation of both these derivatives occurred via a different pathway, and also more slowly, than the less sterically hindered alanine derivative **114b** (***Figure 2.31***).

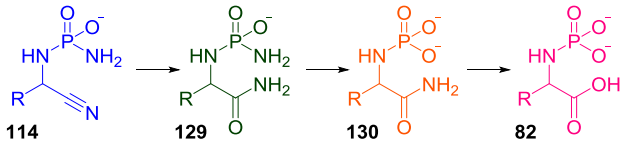


Figure 2.31: *Reaction pathway for the formation of N-phosphoryl amino acid from N-phosphordiamidate aminonitrile for the valine and isoleucine derivatives.*

Valine *N*-phosphordiamidate aminonitrile **114e** underwent complete conversion after 29 days at room temperature. The first intermediate formed was – like the alanine derivative **129b** – the *N*-phosphoramidate aminoamide **129e**, characterised by the presence of a doublet at δ_P 12.06 (J = 10.5 Hz), and a carbonyl carbon shift at δ_C 181.5 (***Figure 2.32***). Unlike the alanine derivative, the next intermediate formed results from the hydrolysis of the phosphordiamidate moiety, rather than from the hydrolysis of the

amide. Its phosphorous peak (δ_P 7.85, d, J = 10.5 Hz) is upfield-shifted, indicating that it is a phosphoryl coupled to just one nitrogen atom; the carbonyl peak is slightly downfield shifted, at δ_C 182.7. This evidence allows us to assign this intermediate as the *N*-phosphono amino amide **130e**. The next species formed is also a monoamido phosphate (δ_P 9.14, d, J = 10.5 Hz) – presumably the *N*-phosphono amino acid **82e**. The carbonyl of the amino acid (δ_C 183.8) is not significantly downfield shifted compared to the amide. Like the reaction of alanine *N*-phosphordiamidate aminonitrile **114b**, the proportion of DAP and MAP in the reaction mixture increased over time, and the unphosphorylated amino acid and amino amide could also be detected (***Graph 2.6***). The proportion of free aminoamide **127e** in the reaction slowly decreased over time, with a concomitant increase in the proportion of amino acid **8e**.

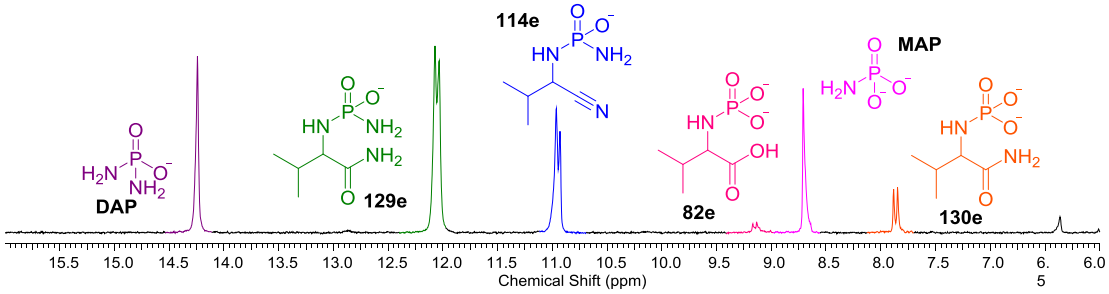
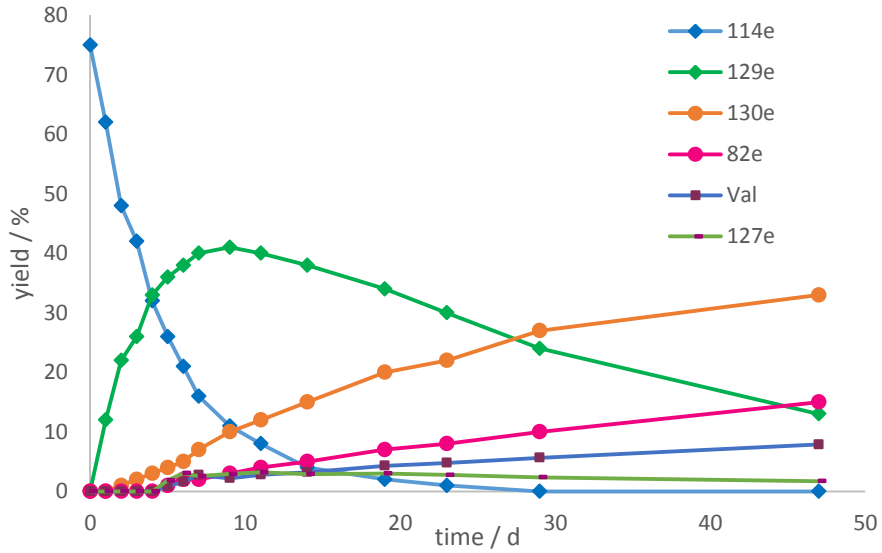


Figure 2.32: *^{1}H-^{31}P coupled NMR spectrum (H_2O:D_2O 1:9, 6.00–16.00 ppm, 284 MHz) to show the reaction of valine phosphordiamidate aminonitrile* ***114e*** *(100mM) at pH 13 at room temperature after 6 d.*



Graph 2.6: *Hydrolysis of purified valine N-phosphordiamidate aminonitrile* ***114e*** *(100 mM) at pH 13 and room temperature. Yields of phosphorylated products calculated by relative integration*

*of identifiable species seen by ^{31}P NMR; the yields of unphosphorylated amino amide and amino acid are calculated by relative integration to **82e** by ^{1}H NMR.*

The reaction proceeded more quickly when heated at 50 °C, but still required 15 days to reach completion. The hydrolysis of crude valine *N*-phosphordiamidate aminonitrile **114e** achieved completion after 18 days, but the unphosphorylated amino acid was again produced alongside the desired product. The reaction of isoleucine *N*-phosphordiamidate aminonitrile **114f** was even slower than that of the valine derivative **114e**. The reaction at 50 °C proceeded more quickly, and also worked well when unpurified **114f** was used.

2.3.6.3 Thiolysis

The fact that the phosphordiamidate moiety of the *N*-phosphorylaminonitriles was apparently incapable of attacking the electron deficient carbon of the nitrile bond led us to investigate activation of this bond by other means so that cyclisation to give an activated peptide bond precursor could take place. It is already known that nitriles can be converted into highly reactive thioamides.

Experiments to understand the behaviour of aminonitriles with hydrogen sulfide first took place over one hundred years ago.[323] Instead of obtaining the corresponding thioamide upon treatment of an aminonitrile with hydrogen sulfide, Johnson and Burnham instead obtained a small amount of the thiodipeptide derivative and a larger amount of the dithiopiperazine: both these products presumably arose from further reactions of the thioamide. Hydrolysis of this cyclic product in concentrated acid resulted in release of the amino acid.[323] It was later found that addition of a ketone to the reaction mixture resulted in the successful formation of a cyclic ketone-adduct that could be hydrolysed to provide an α-aminothioamide.[72,306] We sought to investigate the thiolysis of the nitrile bond of *N*-phosphorylaminonitriles, with the hope that *N*-phosphoryl α-aminothioamides susceptible to intramolecular cyclisation would be produced.

Reaction of purified valine *N*-phosphordiamidate aminonitrile **114e** with sodium hydrosulfide at pH 9 led to the clean formation of the *N*-phosphoryl α-aminothioamide **131e**, with a distinctively downfield-shifted carbonyl resonance at δ_C 212.4. Curiously, analysis by ^{31}P NMR (***Figure 2.33***) suggested that the phosphordiamidate moiety had been converted to a more shielded phosphoryl moiety (ie NH_2 replaced by OH) with an

upfield peak at δ_P 7.43 (d, *J* = 10.5 Hz), but an intermediate in this conversion (either the *N*-phosphordiamidate thioamide **132d** or the *N*-phosphoryl aminonitrile **133d**) could not be detected.

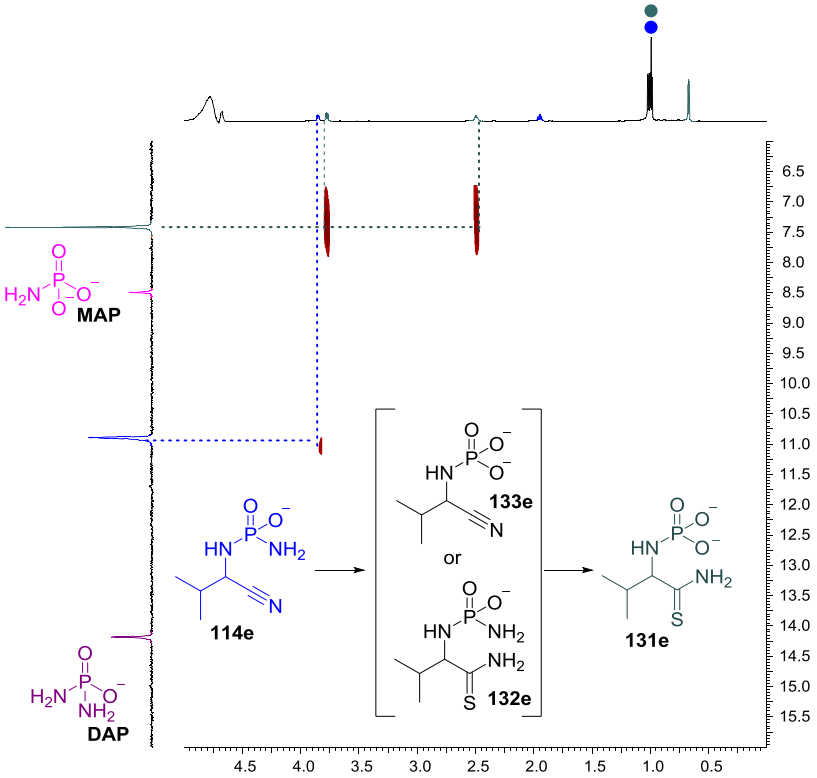
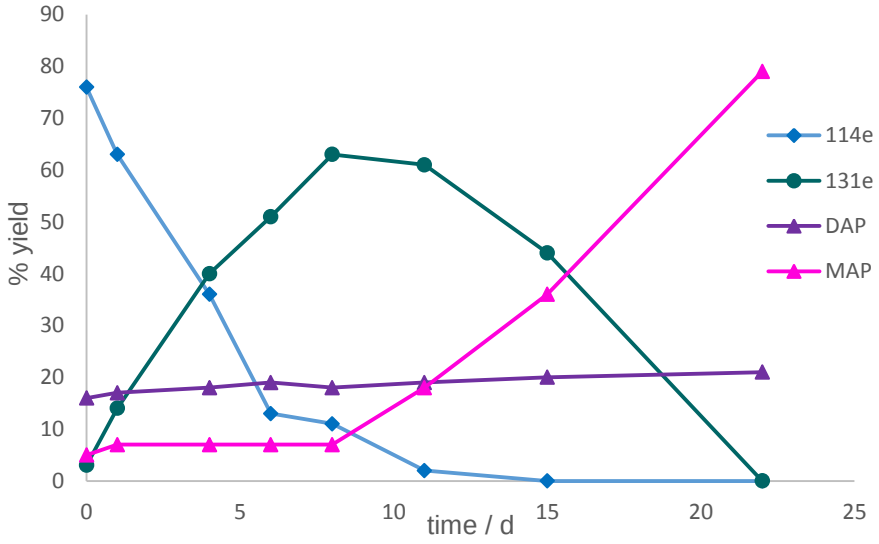


Figure 2.33: *^{31}P- ^{1}H HMBC NMR spectrum (H_2O:D_2O 1:9) showing the reaction of* ***114e*** *(100mM) with NaSH (20 eq) at pH 9 at room temperature after 4 d.*

The *N*-phosphoryl α-aminothioamide **131e** proved to be remarkably sensitive: when left in solution, it decomposed to a complex mixture of unphosphorylated products and MAP (***Graph 2.7***). Attempts to confirm that the phosphordiamidate moiety had indeed converted to a phosphoryl moiety by mass spectrometry were unsuccessful due to the instability of **131e**, with only the unphosphorylated **134e** detected.

This reaction was repeated with a view to determining how long it took to achieve the maximum yield of *N*-phosphoryl α-aminothioamide **131e**. The reaction was therefore not monitored by NMR analysis until day 9, which showed that the unphosphorylated thioamide **134e** had also formed. A similar effect was noticed when unpurified **114e** was used. To allow the yield of the reaction to be accurately calculated, two reactions on the same scale were prepared at once. One was used to follow the progress of the reaction, with aliquots periodically removed for NMR analysis: this underwent clean conversion to the *N*-phosphoryl α-aminothioamide **131e**. The other reaction was left sealed until NMR analysis of the monitored reaction indicated that it would be

complete, at which point it was spiked with pentaerythritol and subjected to NMR analysis. This latter reaction was not clean: the unphosphorylated thioamide **134e** was formed alongside the desired product **131e**, in a 1:2 ratio.



Graph 2.7: *thiolysis of purified valine N-phosphordiamidate aminonitrile* ***114e*** *(100 mM) at room temperature. Yields calculated by relative integration of identifiable species seen by* ^{31}*P NMR.*

A similar phenomenon took place with isoleucine *N*-phosphordiamidate aminonitrile **114f**. Reaction of purified **114f** with sodium hydrosulfide at pH 9 led to the clean formation of the *N*-phosphoryl α-aminothioamide **131f** (again, presence of a shielded phosphoryl moiety visible in the ^{31}P NMR spectrum at δ 7.42 indicated the α-*N*-phosphordiamidate was not present, but an intermediate was not observed). Unpurified **114f** appeared to also undergo clean conversion to a single product – but again only under certain conditions. An undisturbed reaction also formed unphosphorylated thioamide **134f** alongside the desired product (***Figure 2.34***), but only the *N*-phosphoryl α-aminothioamide **131f** was formed when aliquots were regularly withdrawn.

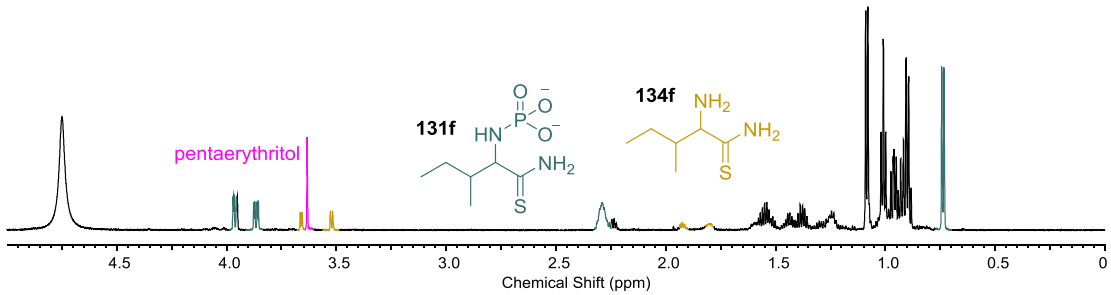
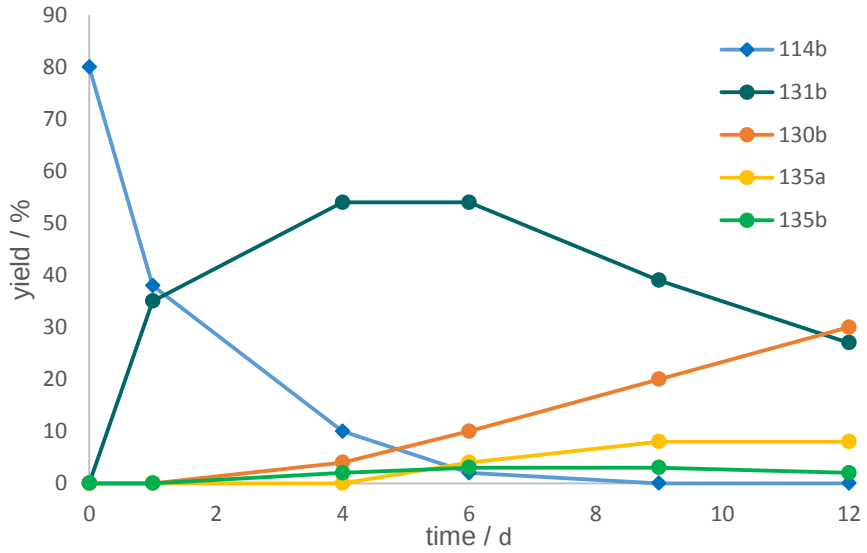


Figure 2.34: *Water-supressed* ^{1}*H NMR spectrum (*H_2O*:*D_2O *1:9, 0.00–5.00 ppm, 700 MHz) showing the undisturbed reaction of unpurified isoleucine phosphordiamidate aminonitrile* ***114f*** *(100mM) with NaSH at pH 9 after 18 d.*

The thiolysis of alanine *N*-phosphordiamidate aminonitrile **114b** resulted in the formation of further products: at first, clean conversion to the *N*-phosphorothioamide **131b** took place, but before the starting material was completely consumed a number of other phosphorylated products started forming (***Graph 2.8***).



Graph 2.8: *thiolysis of purified alanine N-phosphordiamidate aminonitrile* ***114b*** *(100 mM) at room temperature. Yields calculated by relative integration of all species seen by* ^{31}*P NMR.*

The *N*-phosphorothioamide **131b** – like the valine and isoleucine derivatives **131e** and **131f** – contained a phosphoryl moiety (δ_P 7.50). The other species (designated **135a** and **135b**) detected by ^{31}P NMR were also all within the range expected for *N*-phosphoryl groups (δ_P 6.5-9.0), but as they were formed in small amounts within a complex mixture with overlapping peaks in the ^{1}H NMR and ^{13}C NMR spectra, their identity is not clear. The one exception to this was the major by-product, characterised by a ^{31}P resonance at δ_P 8.00 (d, J = 10.5 Hz) and a ^{13}C carbonyl resonance at δ_C 183.7 (***Figure 2.35***). This is in the range expected for an amino acid or amide, and is slightly upfield of the peak assigned as the amino acid carbonyl in the basic hydrolysis of alanine *N*-phosphordiamidate aminonitrile (δ_C 184.8), allowing us to tentatively assign this species as *N*-phosphoryl amino amide **130b**. (Studies on the basic hydrolysis of valine and isoleucine *N*-phosphorothioamides demonstrated that the *N*-phosphordiamidate and *N*-phosphono aminoamide carbonyl peaks are only slightly downfield shifted of the *N*-phosphono amino acids.) Spiking with alanine *N*-phosphoryl amino acid **75b** – formed from the hydrolysis of *N*-phosphordiamidate aminonitrile **114b** at pH 13 – proved that this species was not present among the thiolysis products.

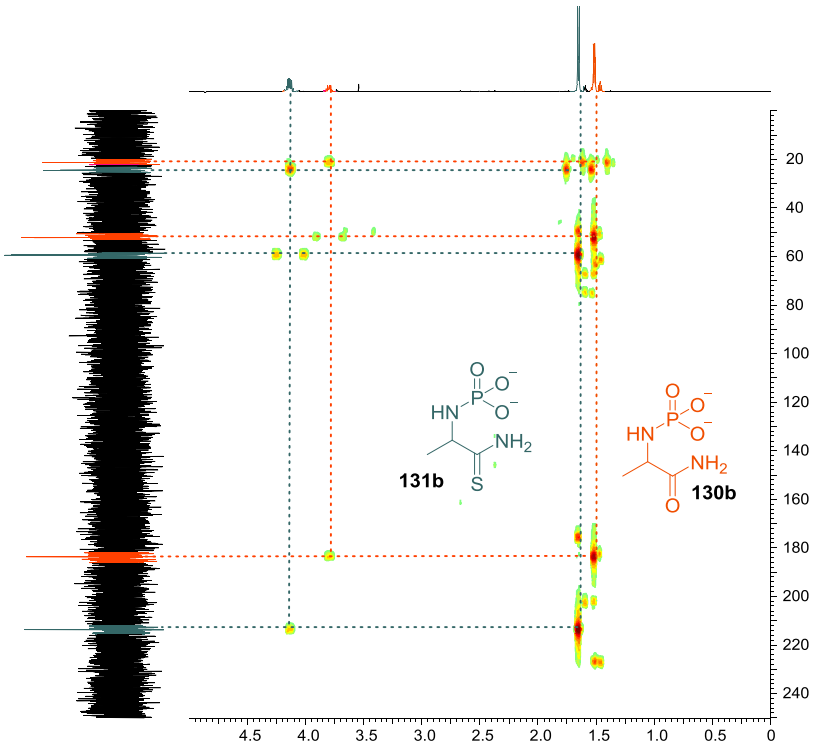


Figure 2.35: *^{13}C-^{1}H NMR HMBC spectrum to show the reaction of unpurified alanine phosphordiamidate aminonitrile (100mM) with NaSH pH 9 after 13 d at room temperature*

It is not clear how this *N*-phosphoryl amide **130b** could be forming: either from direct hydrolysis of the thiocarbonyl, or by an intramolecularly assisted hydrolysis, which would require the formation of a cyclic intermediate such as **136b** (***Figure 2.36***). This cyclic intermediate is very unlikely to be any of the minor phosphorylated products detected by ^{31}P NMR: cyclic phosphate moieties are known to be very downfield shifted (δ 18-26 ppm)[153] and it is likely that such an intermediate is only transiently formed and therefore not detectable by NMR.[76]

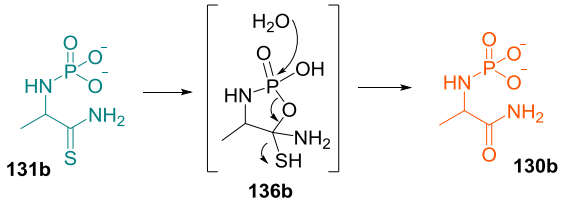


Figure 2.36: *potential conversion of **131b** to **130b** through the intermediacy of cyclic species **136b**.*

2.4 Conclusions and outlook

We have discovered that replacement of ammonia with DAP in the Strecker reaction results in the production of *N*-phosphordiamidate aminonitriles **114** in good yields with high selectivity (***Figure 2.37***). This synthesis has been investigated with a range of aldehydes that are Strecker precursors of proteinogenic amino acids, as well as a number of non-proteinogenic amino acid precursors. We observed (generally) higher yields in the reaction of proteinogenic aldehydes, which could be a factor of the aldehydes investigated (the non-proteinogenic aldehydes used were more hydrophobic and sterically hindered than the natural aldehyde precursors). The successful purification by precipitation has also been demonstrated for selected *N*-phosphordiamidate aminonitriles **114**.

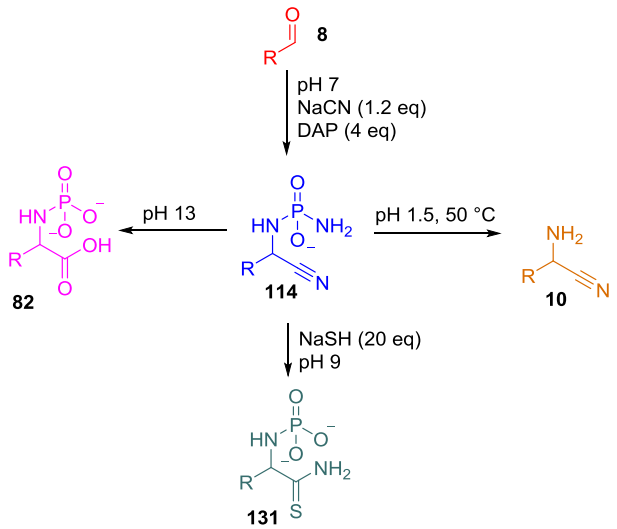


Figure 2.37: *reaction network of N-phosphordiamidate aminonitriles* ***114****, hydrolysing to give aminonitrile* ***10*** *at acidic pH and* ***75*** *at basic pH, and thiolysis to* ***131.***

The reaction shows selectivity towards aldehydes: when mixtures of ketones and aldehydes are used, the α-H *N*-phosphordiamidate aminonitriles **114** are formed in good yield, with the α,α-disubstituted product formed in low yield. It has previously been shown that total selectivity for the synthesis of natural (α-H) aminonitriles **10** can be achieved when a mixture of aldehydes and ketones is exposed to 2-aminothiazole before the addition of Strecker reagents, with selection mediated by crystalline aminal formation and removal of the ketone-containing supernatant.[266] In the absence of 2-aminothiazole, but in the presence of DAP, the phosphoro-Strecker reaction offers a route towards the purely chemical selective formation of proteinogenic amino acid precursors. It is of course possible that the phosphoro-Strecker reaction could take

place using cyanohydrins containing an α-H formed from the reaction of cyanide and aminal. Imprtantly, the entire selection process could therefore take place at pH 7.

Furthermore, it is not clear that total selectivity in the formation of aminonitriles is necessary: the α,α-disubstituted aminonitriles tend to undergo decomposition instead of hydration to α,α-disubstituted aminoamides, but the proteinogenic aminonitriles undergo decomposition to aminoamides (and then hydrolysis to the natural amino acids) instead of decomposition.[313–315] This offers a means of selective formation of the natural amino acids at a later point in their synthesis.

The lower pK_a of DAP compared to ammonia allows the phosphoro-Strecker reaction to take place at neutral pH, unlike the corresponding reaction using ammonia as the amine. The *N*-phosphoramidate aminonitriles **114** are formed in good yield when both ammonia and DAP are present in the reaction mixture at pH 7, with the unphosphorylated aminonitrile formed in low yield. This selectivity is reversed when the reaction is carried out at pH 10.

These observations lead us to propose further explorations of the phosphoro-Strecker reaction to link this reaction to other potentially prebiotic chemical networks. Although our investigations of the phosphoro-Strecker reaction (generally) used aldehydes as starting materials, the reaction of lactonitrile (the cyanohydrin derivative of acetaldehyde) showed that cyanohydrins are also suitable substrates, either with or without the addition of further cyanide, and led to higher yields of the corresponding *N*-phosphordiamidate aminonitriles **114** in some circumstances. Therefore, reaction of DAP with a cyanohydrins formed from the corresponding aminals[266] could take place at pH 7, avoiding the need for the pH of the reaction to be adjusted to 9 before the addition of ammonia. Furthermore, aminal-mediated sequestration of glycolaldehyde and glyceraldehyde (preventing the formation of the more thermodynamically stable dihydroxyacetone from the latter) has been implicated[266] in the stepwise synthesis of ribonucleotides developed by Sutherland,[138] linking the generation of proteinogenic amino acids and ribonucleotides.

The potential for 2-hydroxypentane dinitrile **123**, formed from the cyanosulfidic photoredox elaboration of cyanoacetylene,[64] to be used as an alternative starting material to acrolein **121** in the phosphoro-Strecker synthesis of precursors to glutamic acid **Glu** and glutamine **Gln** has already been discussed. **123** is just one product of a postulated prebiotic cyanosulfidic photoredox network, which also produces valine and

leucine cyanohydrin (**72e** and **72g**) and lactonitrile **72b**,[64] which could be used as starting materials in the phosphoro-Strecker reaction. Notably, **72e** and **72g** are accessed through their α-hydroxythioamide derivatives. Upon irradiation in the presence of hydrogen sulfide and cyanide at pH 7, clean deoxygenation of the α-hydroxythioamide occurs to give the thioamide; continued irradiation in the presence of hydrogen sulfide and cyanide results in the generation of the cyanohydrin. This can be converted into the corresponding aminonitrile in the presence of ammonia, but this reaction requires a more basic reaction mixture (pH 9). Formation of the *N*-phosphordiamidate aminonitrile could instead occur at pH 7, thereby taking place under more similar conditions to the preceding steps. A number of other cyanohydrins, corresponding to aldehydes which have not yet been used as starting materials in the phosphoro-Strecker reaction, are also produced by the same network,[64] offering a range of potential starting materials.

A diverse range of biologically relevant molecules – other than cyanohydrins – can also be produced by the same network, with geology, topology and weather facilitating a proposed spatial and temporal separation of the different reactions and sequential introduction of feedstocks.[64] Therefore, glycolaldehyde and glyceraldehyde – required both for ribonucleotide assemly[138] and to establish a triose glycolysis pathway[154] – are also formed, linking the formation of amino acid precursors to prebiotic informational and metabolic subsystems. A further link between these subsystems is implied by the commonality of DAP (discussed in detail in the introduction of this chapter).

The purified *N*-phosphoramidate aminonitriles **114** can be hydrolysed at acidic pH to afford unphosphorylated aminonitriles. Hydration and hydrolysis of the nitrile bond occurs at high pH, giving the *N*-phosphonoamino acids **75** in good yield. The *N*-phosphonothioamides **131** are formed from the thiolysis of the corresponding *N*-phosphordiamidate aminonitriles when treated with NaSH at pH 9. These *N*-phosphonothioamides could be activated towards peptide synthesis. Similar conversions for these derivatisations can be achieved when unpurified *N*-phosphordiamidate aminonitriles **114** are used: therefore, derivatisation and activation towards peptide synthesis could take place on *N*-phosphordiamidate aminonitriles **114** formed *in situ.* The products formed are not chiral, but it has been shown that mixed-chirality short peptides can promote RNA self-replication;[68] it is therefore possible that chirality was not required for amino acids to fulfil simple functions on the early Earth.

3. Borate as a prebiotic phosphate analogue

3.1 Introduction

Sutherland's seminal synthesis of activated pyrimidine ribonucleotides from simple C2 and C3 molecules via a hybrid sugar-nucleobase scaffold[138] has, despite its elegance, no been without criticism. Chief among the doubts aired over the prebiotic plausibility of this synthesis are: the requisite sequential addition of glycolaldehyde and glyceraldehyde in relatively high purity and a specific order;[243,324,325] the need for highly reactive cyanoacetylene;[326] and requirement for a high concentration of phosphate (up to 1 M) in the required reaction steps of the synthesis. [199,327] Phosphate, due to its low solubility in the presence of divalent cations,[180] has perhaps received the most criticism – but phosphate fulfils several essential roles required during nucleotide synthesis.

The prebiotic availability of phosphate on the early Earth is not clear. The most important naturally occurring phosphate minerals on Earth are only slightly soluble in water:[328] models predict that as little as <0.1 M phosphate would be present in the early Earth's oceans.[329] Evaporating seawater would not result in higher concentrations of phosphate in solution, as it has been proposed that the high natural abundance of calcium (the fifth most abundant element in the Earth's crust[199]) would lead to the precipitation of phosphate as the calcium-containing mineral, apatite.[199,328] Furthermore, though polyphosphates are more soluble, they are not thought to be geologically significant on the early Earth.[182] Heating apatite can, however, result in the volatilisation of condensed phosphates (including pyrophosphate and tripolyphosphate)[328], albeit in low yield.[182] Moreover, the mineral schreibersite (which could have arrived on Earth from extra-terrestrial sources) is more soluble in water than other more common terrestrial phosphate-containing minerals;[328] and calcium phosphite, a product of schreibersite corrosion, is around a thousand times more soluble than apatite in water.[183] These considerations have led some researchers to propose schreibersite and its phosphite corrosion products as an important prebiotic phosphorous source.[2,184,328]

Nevertheless, it appears prudent to explore adaptions to this synthesis that place less emphasis on the need for phosphate. Sutherland's route to activated pyrimidine nucleotides has been noted for its inclusion of phosphate from the start of the synthesis[57,119,329] (acting as a general acid-base catalyst, a pH and chemical buffer, and finally as a reagent for the key rearrangement step to deliver the correct sugar

stereochemistry and form an activated 2′,3′-cyclic phosphate), thereby removing an environmental constraint to introduce (or remove) phosphate at any particular stage. The key phosphorylation is displays a parsimonious use of reagents: phosphorylation can occur either when the anhydronucleoside is heated with pyrophosphate (a by-product of the preceding step) in urea (a by-product of the first step of the synthesis), or when heated with ammonium dihydrogen phosphate and urea in formamide. Both these methods, however, require a drying step to transition from the aqueous phosphate-buffered conditions used in previous steps.

An alternative reagent for the rearrangement of the anhydronucleoside that could be used in aqueous solution would offer a useful adaption to this synthetic route. Ideally, this reagent should be one that is already implicated in the selection and reaction of potentially important bio-molecule precursors. One such reagent is borate.

3.1.1 Borate in prebiotic chemistry

Boron as an uncomplexed element is not naturally found on Earth[330,331], although boron containing minerals on Earth are known, albeit relatively rare[161]; conversely, no boron minerals have been reported in meteorties.[332] The boron found on Earth nearly always bound to oxygen,[331] and is unevenly distributed, being found mostly in the crust (boron is cycled between the surface and shallow areas of the crust[159]), and even there it is unevenly distributed.[332] Tourmaline (a borosilicate) can be found in rocks formed 3.8 Gya, implying that aqueous borate was present in the ocean and sunk into sediments at that time;[159] and boron minerals and borate-rich waters may have also been present before this point in time.[332] It has been suggested that, due to the absence of continental weathering and other factors which cycle boron today, the cycling of boron in the prebiotic world was controlled only by submarine hydrothermal activites.[333] Although the earliest geological evidence of boron minerals is only 3.8 Ga old, this absence of an earlier record does not exclude the presence of boron minerals and borate-rich waters before this point in time.[332] Although the prebiotic availability of borate is unclear, it has been proposed to play a useful role in some potentially prebiotically important reactions.

Borates and boronic acids are known to reversibly form cyclic esters with *cis*-diols,[334] and therefore can be viewed as analogues of 2′,3′-cyclic phosphates. The stability of the borate-diol ester depends on the specific structure of the diol, as well as the

concentration and ionisation constant of the participating molecules.[161] Cyclic compounds, with the *cis*-diol orientated to point in the same direction, towards borate, are particularly well-suited to co-ordination.[90]

The RNA world theory implicates ribose at the origin of life, but ribose is inherently unstable at pHs near neutral.[332] Ribose, containing a *cis*-diol, is known to reversibly form cyclic esters with borates and boronic acids[334] – in fact, ribose is one of the best chelators of boric acid.[173] The resulting cyclic ribose-borate complex is stabilised, and prevents the polymerisation and thermal degradation of ribose.[173] The furanose form of ribose (found in nucleosides) forms a stronger complex with borate than the pyranose form,[161] due to a network of hydrogen bonds.[90] Although the stability of all pentoses increases with an increasing concentration of borate, ribose shows the greatest increase in stability and has the highest binding affinity with borate (which may be responsible for its significant increase in stability).[159] As ribose is the least stable of the pentoses, however, this co-ordination to borate does not necessarily mean that the borate-ribose complex is more stable than other borate-pentose complexes. Nevertheless, this stabilisation by borate may have offered a potential (pre-)evolutionary advantage to ribose-containing nucleic acids.[173,161,160]

Boron may also be implicated in the formation of ribose, and other pentose sugars, to prevent their decomposition in the alkaline conditions required for their formation in the formose reaction (***Figure 3.1***). The formose reaction requires the nucleophilic addition of the enolate form of glycolaldehyde to formaldehyde; the presence of borate is reported to not affect this step. The glyceraldehyde formed, however, contains a 1,2 *cis*-diol and so forms an anionic complex with borate: the negative charge has been proposed to inibit enolisation and so the complex cannot act as a nucleophile. Complexation also prevents isomerisation to dihydroxyacetone (which, in fact, appears to be the key nucleophile in extension of formose homologues past the triose sugars).[149] The glyceraldehyde-borate complex can, however, still react as an electrophile with glycolaldehyde to give the pentose sugars. As the cyclic forms of the pentoses do not have a carbonyl moiety, they form more stable, less reactive complexes when chelated by borate. Borate-pentose complexes are therefore stable for days under the conditions of the formose reaction, compared to minutes in the absence of borate.[147] Borate can complex to two pairs of hydroxyl groups, so at low

concentrations borate preferentially forms diester complexes by cross-linking two *cis*-diols.[161]

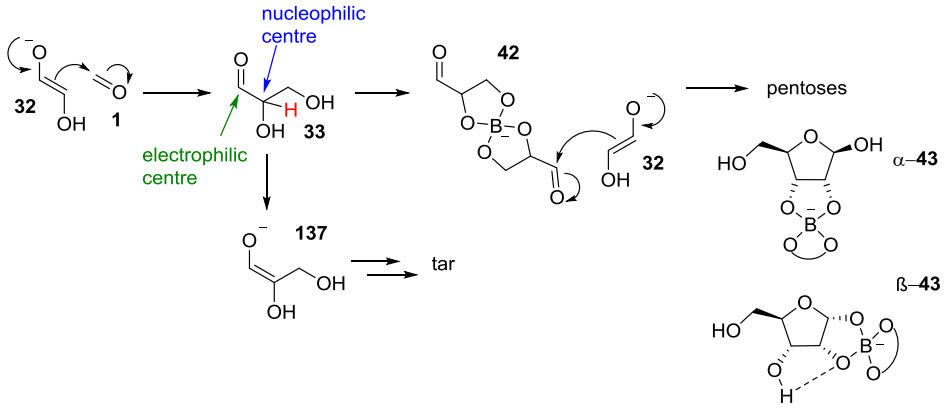
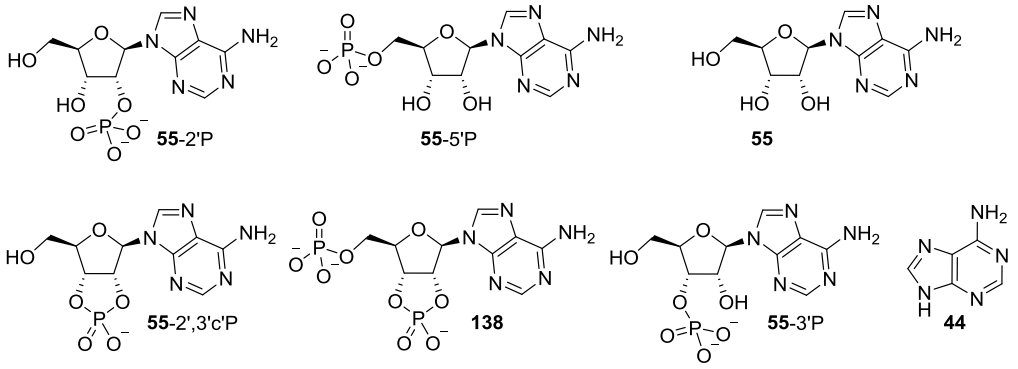


Figure 3.1: *Formation of pentoses in the presence of borate. The reaction of formaldehyde* ***1*** *and glycolaldehyde* ***32*** *yields glyceraldehyde* ***33****. The electrophilic carbon atom (green), nucleophilic carbon atom (blue) and the acidic proton (red) in* ***33*** *are highlighted.* ***33*** *can enolise to* ***137****, which is capable of reacting as a nucleophile, leading to a complex mixture of products.* ***33*** *can coordinate to borate:* ***42*** *still contains an electrophilic centre (and so can undergo attack by* ***32****) its nucelophilicity is supressed by the co-ordination of borate. The cyclic forms of the pentoses formed co-ordinate borate (two ribo-isomers, α-****43*** *and ß-****43,*** *are shown).*

Unfortunately, the authors of this study do not provide yields:[147] the C4 and C5 sugars are still formed in a complex mixture, and yields for some of these products – but interestingly not ribose – are reported in a later paper.[148] Only around 50% of the formaldehyde present at the start of the reaction, however, can be accounted for:[148] it is likely the remainder undergoes an unproductive Cannizzaro disproportionation to give methanol and formate. It has been suggested that the stabilisation afforded to ribose by co-ordination with borate under the conditions of the formose reaction is modest, with ribulose being more significantly stabilised.[119] Furthermore, an excess of glycolaldehyde is required to overcome the inhibition provided by borate; otherwise, the formation of branched pentoses is observed.[148]

This ability of boronic acids to form esters with *cis*-diols has also been exploited to form borate-complexed nucleotides and nucleosides: the borate-esterification of several nucleotides has been investigated.[335,336,337] It was found that the borate esterification of NAD and NADH is pH dependant, with maximum esterification under alkaline conditions, although co-ordination of borate to the 2′,3′-*cis* diol remains significant at pH 7.[336] In fact, both borate and boric acid can co-ordinate to the 2′,3′-*cis* diols of these

nucleosides,[336] although *cis*-diol containing molecules react with borate to form borate linkages that are more stable to hydrolysis than the trigonal borate esters formed from than boric acid.[161] The presence of a 5′-phosphate group does not prevent borate co-ordination to the 2′,3′ diol of nucleotides.[337] The chemistry of boron co-ordination to nucleotides and nucleic acids has been widely investigated in boro-analogues of nucleic acids, in which phosphate linkages are replaced by borate groups.[334,338,339]



	55-5'P	**55**	**55**-2',3'cP	**138**	**55**-3'P	**44**
borate	29%	65%	–	–	–	–
no borate	15%	18%	13%	9%	10%	4%

Figure 3.2: *Rearrangement of adenosine-2′-phosphate **55**-2'P when heated with phosphate in urea.*[175] *In the presence of borate, the only phosphorylated product is **55**-2'P, which presumably forms because borate co-ordinates the cis-diol and directs phosphorylation to the free 5′-OH. A complex mixture of phosphorylated and un-phosphorylated products are formed in the absence of borate.*

Incubation of adenosine 2′-phosphate **55**-2'P with urea and phosphate results in isomerisation to the corresponding 5′-phosphate **55**-5'P in the presence of borate (***Figure 3.2***).[175] Although a mechanism for this rearrangement is not clear, it appears likely that the reaction begins with reversible dissociative transfer of phosphate between urea and the nucleotide hydroxyl groups, whilst borate co-ordinates (once free) the *cis*-diol of the nucleotide and therefore directs phosphorylation to the 5′-OH. Incubation of adenosine 2′-phosphate with urea in the absence of borate resulted in a complex mixture of phosphorylated adenosines, but favours the irreversible formation of adenosine 2′,3′-cyclic phosphate **55**-2',3'cP.[175] It appears likely that prolonged incubation of nucleotides and phosphate, given the reversible nature of the 2′,3′-borate moiety and the irreversible nature of the 2′,3′-cyclic phosphate moiety, would ultimately

also furnish nucleotide 2′,3′-cyclic phosphates – but at early time points during incubation borate has a pronounced effect.

Uridine *ribo*-**60** has been observed to extract borate (by complexation with its 2′,3′ diol) from luneburgite (a mineral containing magnesium, borate and phosphate), with concomitant release of phosphate in aqueous solution.[199] In pure water, the amount of phosphate released from luneburgite is minimal.[199] When uridine is heated with luneburgite in urea, uridine-5′-phosphate is obtained (albeit in only 15% yield); when orthophosphate is substituted for luneburgite, a complex mixture of phosphorylation products are produced, further showing the potential of complexation with borate in controlling the reactivity of the *cis* hydroxyl moiety.

Boron has also been suggested to have a role in other areas of origin of life chemistry. Borate has been proposed to assist the polymerisation of glycine in water by forming a complex with the carboxylic moieties of two amino acids.[173] Formamide polymerises in the presence of borates: in the absence of borate, purine is the only recovered product (and, indeed, is the main reaction product irrespective of the boron mineral); in the presence of different types of borate mineral, lactic and pyruvic acids, adenine and pyrimidine nucleobases are also synthesised (albeit in yields <0.1%).[340]

It has been proposed that borate could also act as an activating and leaving group for sugar glycosidation, and although a borate-ribose-adenosine complex has been detected by mass spectrometry following the addition of adenosine to an aqueous borate-ribose solution, details of yield, connectivity and the stereochemistry of the species formed have yet to be reported.[17] It is implied that complexation of ribose with boron salts aids purination: the proposed reaction scheme resembles the formation of nucleosides in present-day cells, with borate the prebiotic equivalent of today's pyrophosphate,[173] but this remains to be demonstrated. It is also possible, and consistent with the mass spectroscopy results, that borate co-ordinates to the ribose diol whilst the exo-cyclic amine of adenine forms a reversible aminal with ribose.

3.2 Aims

The aim of this project was to investigate the potential of borate as a prebiotically plausible phosphate analogue. Although borate is well known for its ability to co-ordinate *cis*-diols, it was hypothesised that *ara*-**63** could reversibly form a complex with

borate at its 3′-OH. It is known that the vacant p-orbital of boron (in its trigonal planar form, in which it is an isoelectronic neutral analogue of a carbocation) can readily accept a pair of electrons (ie from 3′-OH of *ara*-**63**) to form an anionic tetra-coordinated sp^3 species.[161] The sp^3 species formed is chemically stable and could then act as a nucleophile.[161] Intramolecular nucleophilic attack of the 3′-OH co-ordinated borate on the 2′-carbon atom would result in stereochemical inversion of this 2′-position, to form **33**, an analogue of 2′,3′-cyclic cytidine monophosphate *ribo*-**29**-2',3'cP **(*Figure 3.3*)**. The reaction mechanism is essentially similar to the formation of the phosphorylated analogue: [138] reversible group transfer followed by intramolecular rearrangement. The borate could then be used to control the reactivity of the 2′,3′-hydroxyls and direct site-specific 5′ phosphorylation delivering *ribo*-**29**-5'P [175,199] (or, indeed, *ara*-**63-5′P** could be used[320]), or could later be removed and replaced with a phosphate group at the same position.

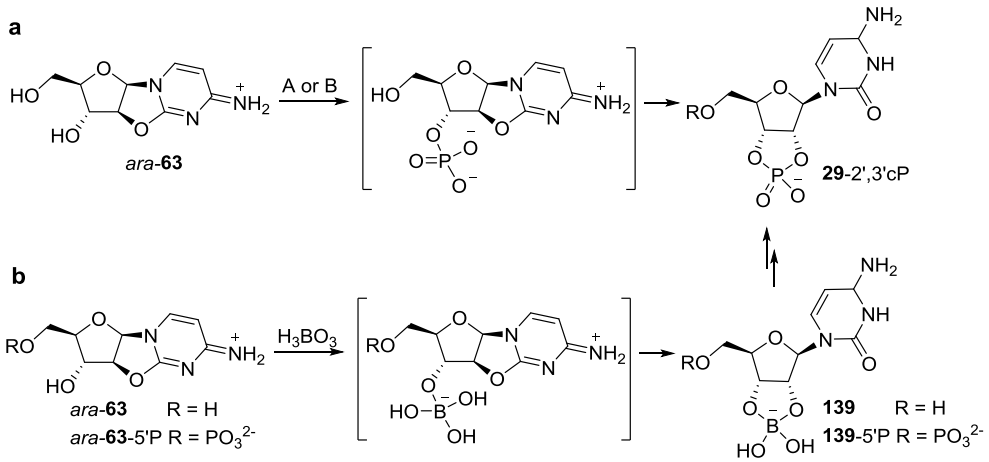


Figure 3.3a: *Reported synthesis of ribo-**29**-2',3'cP by urea-mediated phosphorylation of anhydrocytidine ara-**63**.*[138] *(A: PPi, urea, heat; B: Pi, urea, formamide solution, heat)*

***b**: Proposed route from anhydrocytidine ara-**63** to cytidine-borate complex **139** or **139**-5'P by borate-mediated stereochemical inversion. (R = H or PO_3^{2-})*

3.3 Results and Discussion

3.3.1 Borate as a phosphate analogue in pyrimidine synthesis

An initial investigation into the ability of borate to act as a prebiotic phosphate analogue in the rearrangement of ancitabine *ara*-**63** was carried in aqueous solution out using boric acid at pH 6. The pKa of boric acid is 9.2,[337] so at pH <9 it exists in solution primarily as trigonal planar boric acid and can act as an electrophile. We hypothesised this could be available to co-ordinate to the 3′-OH of *ara*-**63**. At room temperature, using 3.5 eq of boric acid, ^{1}H NMR analysis of this reaction revealed very little change

over time: after 12 days, just 5% of *ara*-**63** had converted to arabinose cytosine *ara*-**29**. This reaction was further investigated at a range of pH values and with a greater proportion of boric acid (6.5 eq.), which only demonstrated that hydrolysis was faster at higher initial pH values (***Table 3.1***). A control reaction (without boric acid) also showed the formation of *ara*-**29**, albeit with only 2% conversion after 10 days at pH 6.3.

	Conversion to arabinose cytosine (%)				
time / d	**pH 6.9**	**pH 6.4**	**pH 5.8**	**pH 5.6**	**pH 4.5**
0	0	0	0	0	0
1	3	2	1	0	0
4	10	4	2	1	0
5	13	5	2	2	0
7	-	15	11	5	2
11	46	21	12	6	3

Table 3.1: *Reaction of ancitabine ara-**63** and boric acid (6.5 eq) at room temperature.*

The identity of arabinose cytosine *ara*-**29** was confirmed through NMR spiking and comparison studies. It was clear from spiking the ^{1}H NMR (***Figure 3.5***) that the structure formed was not in the same conformation as cytidine *ribo*-**31** (and therefore was not the intramolecular nucleophilic substitution product **139**). Instead, the structure formed was in the same conformation as *ara*-**29**, and therefore plausibly the hydrolysis product **140**, or other borate-complexed derivatives of this.

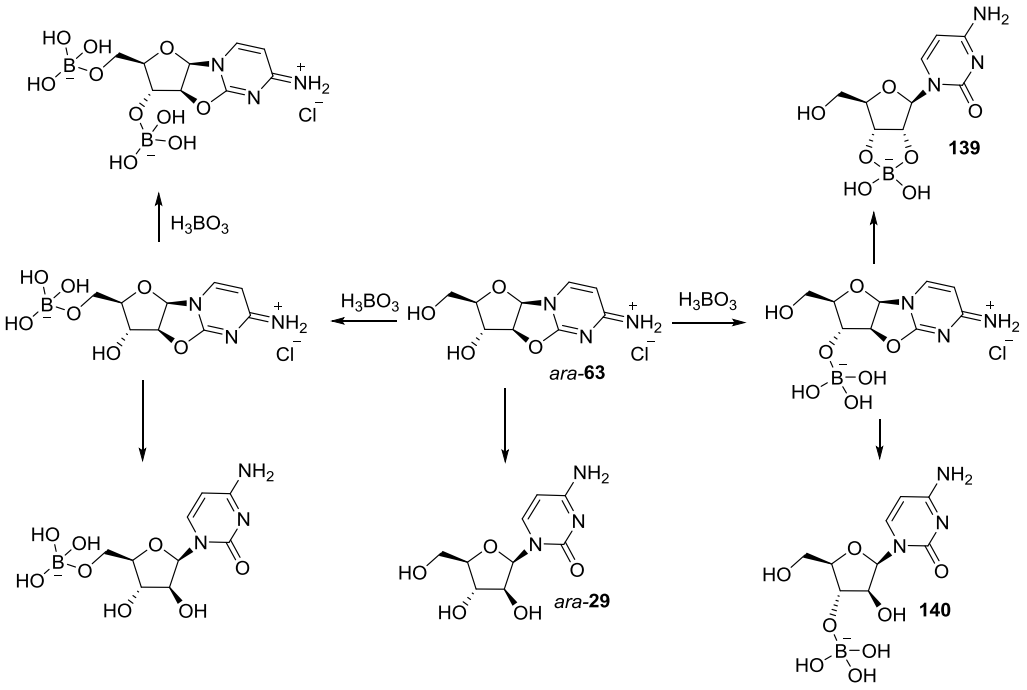


Figure 3.4: *Potential intermediates and products for the reaction of ancitabine ara-**63** with boric acid.*

Borate-complexed ^{13}C chemical shifts are distinctively deshielded.[341] However, it was thought that if the boric acid present in solution was capable of forming a complex with arabinose cytosine *ara*-**29** generated during the reaction, it would also be capable of forming a complex with *ara*-**29** used to spike the sample. NMR spiking experiments therefore could not conclusively prove the identity of the species formed during the reaction, but it was obvious they were derived from 2,2′-hydrolysis.

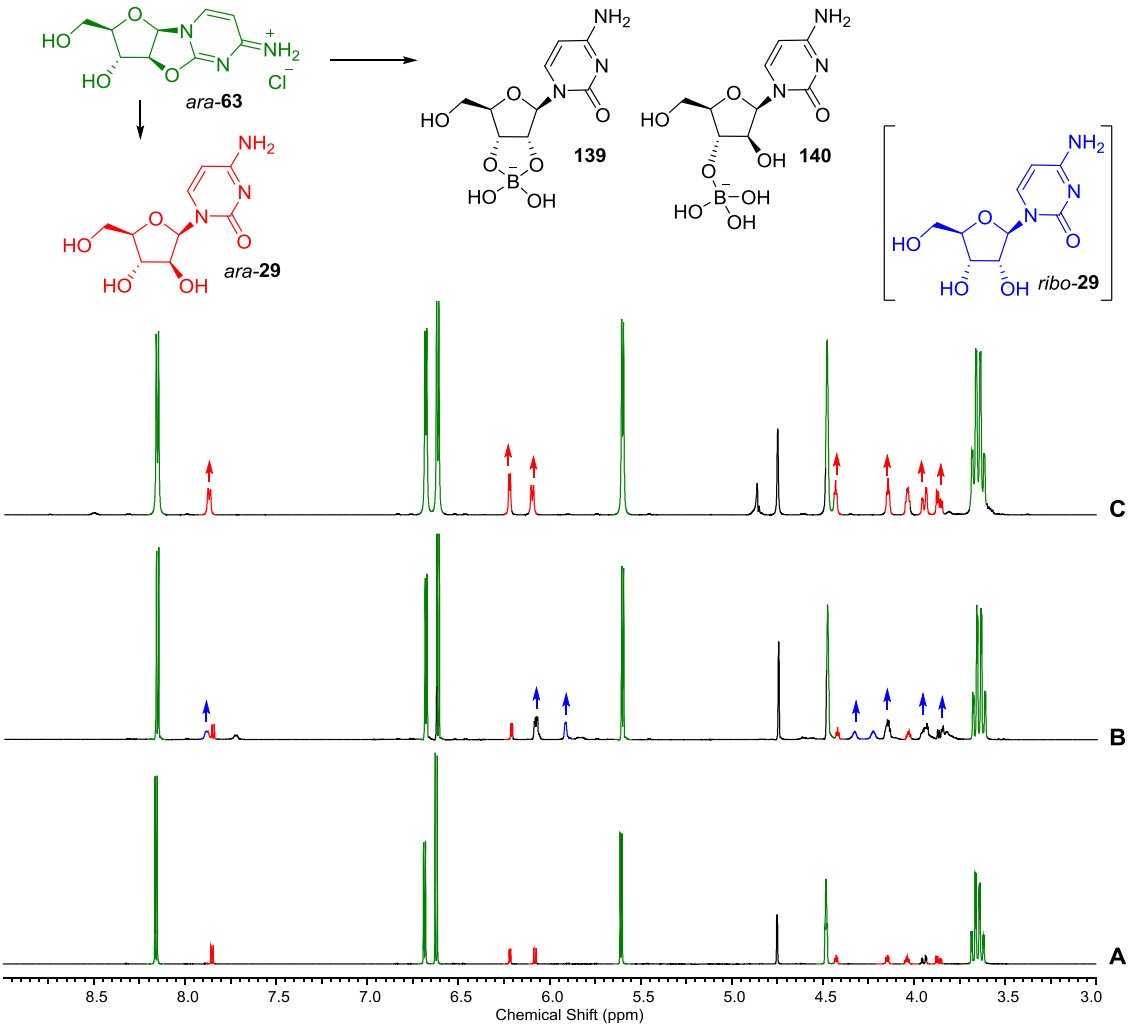


Figure 3.5: *Determining the conformation of the species formed through NMR spiking studies. ^{1}H NMR spectra (600 MHz, 3.0-9.0 ppm, D_2O) of:* ***A****) the reaction mixture of ancitabine ara-****63*** *with $B(OH)_3$; and separately spiked with* ***B****) cytidine ribo-****29****; and* ***C****) arabinose cytosine ara-****29****.*

Therefore, the ^{13}C NMR spectra of the reaction mixture and *ara*-**29** at the same pH were obtained separately. It can be seen from these that the arabinose cytosine-like structure formed in the reaction mixture and authenticated arabinose cytidine are the exactly the same: no borate-nucleoside complexes were formed (***Figure 3.6***). *ara*-**63**

is therefore simply undergoing hydrolysis (a well-known degradation pathway[342]) rather than the desired borate complexation. No ribose isomer was observed.

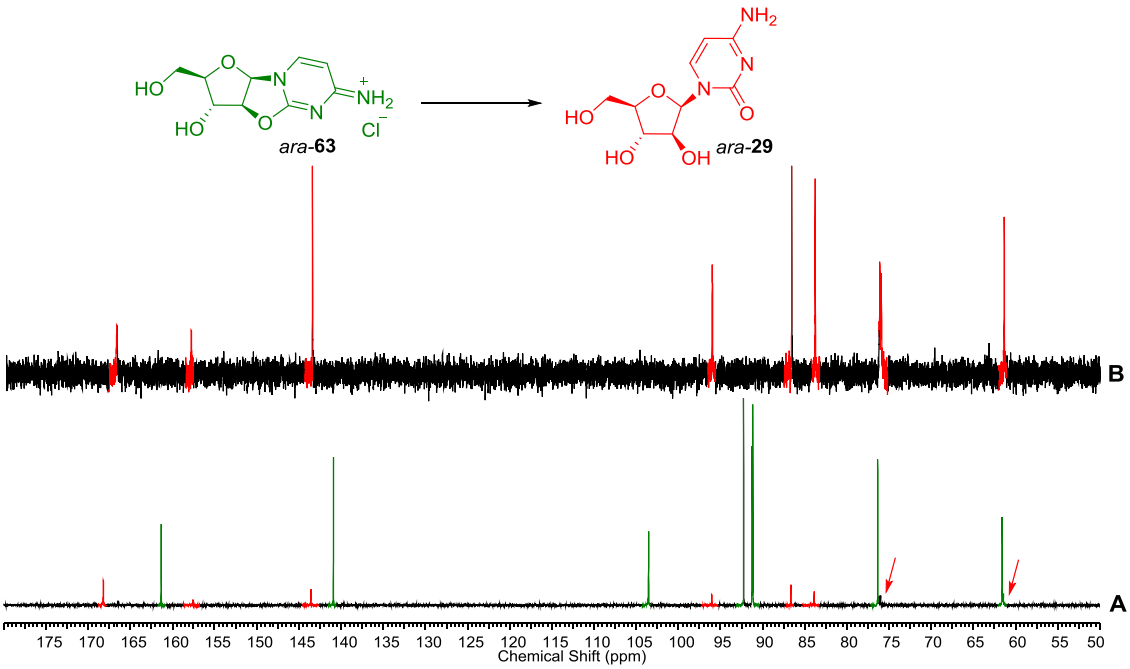
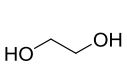
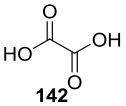


***Figure 3.6**: ^{13}C NMR spectra (150 MHz, 50-180 ppm, D_2O) of: **A**) the reaction mixture of ancitabine ara-**63** with $B(OH)_3$; and **B**) commercial arabinose cytosine ara-**29**.*

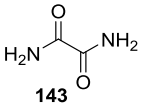
Nevertheless, further changes were made to the reaction conditions in an attempt to promote borate-nucleoside co-ordination. Conversion proved to be much faster when the reaction was heated. Systematic evaluation of the effect of initial pH value on the reaction again revealed that greater conversion to arabinose cytosine *ara*-**29** occurred at higher pH values (for example, ***Table 3.2*** entry 3 compared to entry 7). Analysis of ^{1}H NMR spectra showed that *ara*-**29** was forming both in the presence of boric acid, and in control reactions with no borate (***Table 3.2***, entries 1 and 2), but that conversion was much slower in the control reactions. It could be supposed that this was because boric acid was catalysing the hydrolysis of ancitabine *ara*-**63**: however, careful inspection of the initial and final pH values reveals that pH rather than borate is likely to be the major influencing factor in the rate of conversion. Boric acid acts as a pH buffer, maintaining the pH near neutral; without buffer, the pH falls and less conversion occurs. There is a clear relationship between final reaction pH and conversion to *ara*-**29**.



141



142



143

Entry	*ara*-**63** (mmol)	H_3BO_3 (eq.)	additive (2.3 eq.)	initial pH	Heat time (h)	final pH	*ara*-**63** (% yield)	*ara*-**29** (% yield)
1	0.04			5.5	20	3.8	95	5
2	0.03			6.4	22	3.7	98	3
3	0.06	10.8		6.9	20	5.7	23	77
4	0.06	10.8		6.5	20	5.3	50	50
5	0.06	10.8		6.4	20	5.2	59	41
6	0.06	10.8		6	20	5.1	89	11
7	0.06	10.8		5.5	20	4.7	71	29
8	0.06	10.8	**141** (2.3 eq.)	6.9	20	5.6	32	68
9	0.06	10.8	**142** (2.5 eq.)	6.7	20	5.4	73	27
10	0.06	10.8	**143** (1.8 eq.)	7	20	6.1	22	78

Table 3.2: *reaction of ancitabine ara-**63** and boric acid (10.8 eq) at 50 °C. Structures of ethylene glycol **141**, oxalic acid **142** and oxamide **143** are shown above.*

It was hypothesised that the presence of a diol additive would co-ordinate the borate, locking it in anionic tetra-co-ordinated sp^3 form, which is able to act as a nucleophile. However, the addition of ethylene glycol **141** and oxamide **143** (***Table 3.2***, entries 8,10) did not alter the nature of the species formed. The reaction with oxalic acid **142** (***Table 3.2***, entry 9) resulted in relatively little conversion; however, the solution drops from pH 6.7 to 5.4 over 20 h in the presence of **142**, and this drop in pH decreases the rate of anhydronucleoside hydrolysis.

The reaction of *ara*-**63** and other boron complexes were also investigated. Borax, like boric acid, proved to be a disappointing reagent: instead of the desired co-ordination, hydrolysis to arabinose cytosine *ara*-**29** was observed. With sodium perborate, however, an unexpected reaction took place, with two major products formed. Spiking experiments proved that one of the products was *ara*-**29** (11% conversion after 44 h at 50 °C); but that cytidine *ribo*-**31**, uridine *ribo*-**60**, arabinose uracil *ara*-**60**, and 2,2′-cyclo uridine **144** were not present. Interestingly, the other major product was arabinose amino-oxazoline *ara*-**62** (11%). A minor amount of arabinose oxazolidinone *ara*-**64** (1%) also formed, alongside several small by-products observed, but not identified, in

the ^{1}H NMR (***Figure 3.7***; ^{1}H NMR spectra showing the spiking studies can be seen in ***Figures 8.3.1*** and ***8.3.2*** in the experimental section for this chapter).

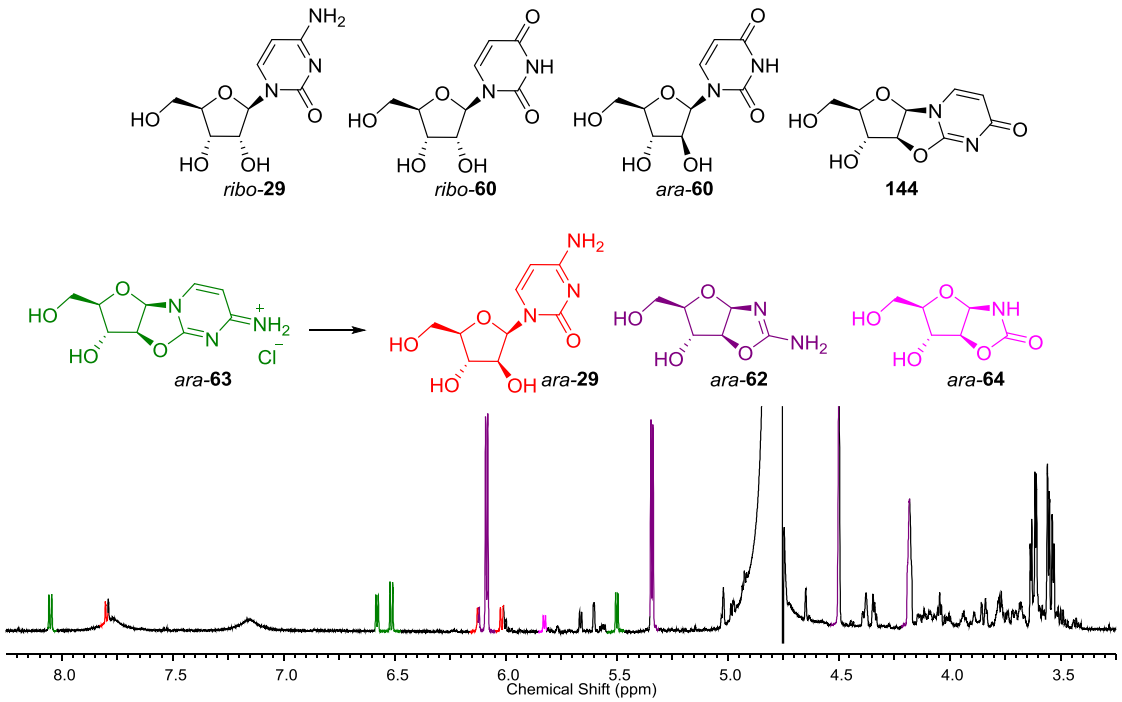


Figure 3.7: *^{1}H NMR spectrum of the reaction mixture of ancitabine ara-**29** with $NaBO_3$ (5.7 eq) after 20 h at 60 °C.*

It is likely that these other unidentified species are intermediates formed during the synthesis of arabinose aminooxazoline *ara*-**62** and arabinose oxazolidinone *ara*-**64** from ancitabine *ara*-**63**. It is not clear how this decomposition is occurring; however, it is known that sodium perborate decomposes to give borate and hydrogen peroxide in aqueous solution,[343] and that sodium perborate has been used to effect a variety of functional group oxidations.[344] It is likely that hydrogen peroxide undergoes Michael addition at the C6 carbon atom of *ara*-**63**, which then leads to nucleobase fragmentation. A tentative mechanism for this reaction is shown in ***Figure 3.8***, but the 3-carbon fragment was not observed in our experiments and further investigation of this mechanism was not pursued.

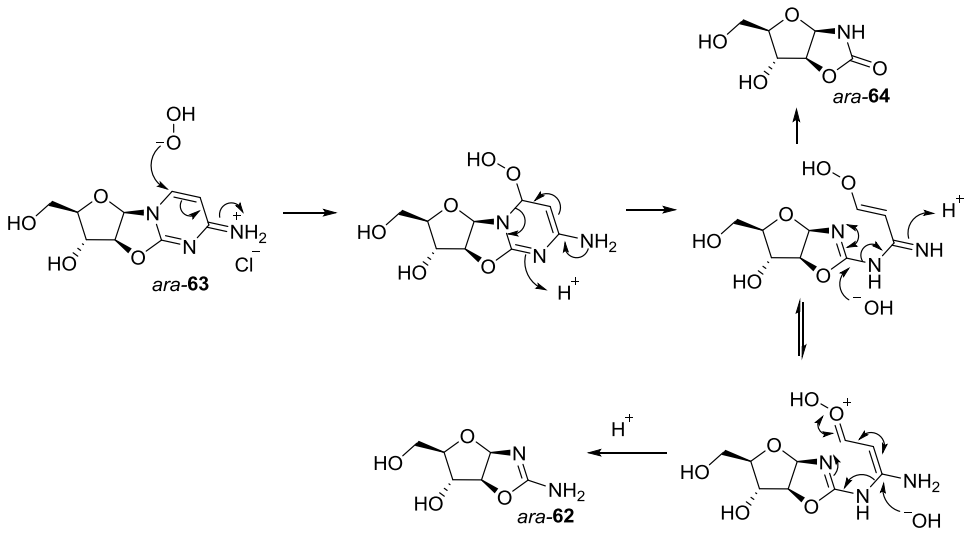


Figure 3.8: *Possible mechanism for the formation of arabinose amino-oxazoline ara-**62** and arabinose oxazolidinone ara-**64** from ancitabine ara-**63** in the presence of sodium perborate.*

Conversion again was found to have a positive correlation with increasing initial pH values (although the pH of the solution fell significantly over the course of the reaction), and with reaction temperature. Various diol additives (***Table 3.3*** entries 7-9) were again introduced in an attempt to chelate borate and fix it in a conformation in which it could easily co-ordinate to a nucleoside diol, but again these had no effect on the nature of the products of the reaction. The reaction in the presence of oxalic acid **142** (entry 8) produced the highest proportion of arabinose amino-oxazoline *ara*-**62**; this is likely due to the buffering effect on reaction pH.

	ara-**63** (mmol)	$NaBO_3$ (eq.)	Add. (eq.)	initial pH	Heat time (h)	final pH	% yield				
							ara-**63**	*ara*-**29**	*ara*-**62**	*ara*-**64**	Unknown
1	0.05	2.8		7.5	22	3.9	56	5	28	<1	10
2	0.05	2.8		7.0	22	3.6	75	4	13	<1	7
3	0.05	2.8		6.5	22	3.6	83	3	13	<1	<1
4	0.05	2.8		6.3	22	3.4	82	4	10	<1	3
5	0.05	2.8		5.9	22	3.3	81	4	11	<1	3
6	0.05	2.8		5.4	22	3.0	89	4	5	<1	1
7	0.05	2.8	**141** (2.8)	6.8	22	3.8	77	3	20	0	<1
8	0.05	2.8	**142** (1.8)	6.7	22	4.6	47	7	35	2	9
9	0.05	2.8	**143** (2.0)	6.8	22	3.8	70	2	24	0	4

Table 3.3*: reaction of ancitabine and sodium perborate (2.8 eq) at 50 °C. Additives: ethylene glycol **141**, oxalic acid **142** and oxamide **143**.*

3.3.2 Borate as a phosphate analogue in purine synthesis

As ancitabine *ara*-**63** was shown to be disappointingly sensitive to hydrolysis, it was decided to attempt a borate-induced rearrangement and co-ordination on a different anhydro-nucleoside. Whereas *ara*-**63** readily hydrolyses in aqueous solution at near-neutral pH, the 2′,8-anhydro purines are known to be remarkably resistant to hydrolysis,[342,345] a result of the exceptionally stable 2′,8-anhydro bond.[345] Even upon extended incubation at high temperature and pH, 2′,8-anhydro adenosine **145** has not been observed to undergo hydrolysis. At pH 11, 2′,8-anhydro adenosine isomerises to form 5′,8-anhydro arabinoadenosine *ara*-**146**,[342] and at pH 13 the 8-oxo-2′,3′-expoxide **147** forms.[342,345] The conversion of 2′,8-anhydro adenosine into 2′,3′-epoxy-8-oxo-adenosine is reversible.[345]

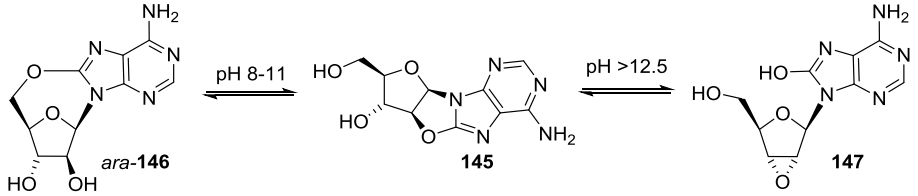


Figure 3.9: *Isomerisation, not hydrolysis, of 2′,8-anhydro adenosine takes place in alkaline aqueous solution.*

The reaction of 2′,8-anydroguanosine **148** and borate was therefore also investigated, using both boric acid and sodium perborate at a range of pH values and temperatures. **148** was found to be stable in aqueous solution at neutral pH with both boric acid and sodium perborate at room temperature. When heated in the presence of either borate source, **148** was found to undergo slow conversion to another nucleoside species at pH <8, and more rapid conversion to the same species at pH >8.5. A second nucleoside species was also observed to form at more alkaline pH. The same two species also formed in a control reaction, without the presence of a borate source, when heated at pH 9.8.

Although literature NMR data for the expected products of the isomerisation of 2′,8-anydro-guanosine – 5′,8-anhydro arabinoguanosine *ara*-**149** and 2′,3′-epoxy-8-oxo-guanosine **150** – are not available, comparison of NMR data available for the adenosine analogues suggests these are indeed the products, and that no complexation with borate was taking place. The product formed predominantly at lower pH is believed to be 5′,8-anhydro-arabinoguanosine *ara*-**149**. It is characterised by the presence of a doublet anomeric peak (δ 6.29 ppm, J = 6.5 Hz in D_2O, compared to δ

6.30 ppm, d, J = 6.4 Hz for the adenosine analogue[342]), and a pair of well-separated, deshielded H-5′ peaks (at δ 4.56 ppm and δ 4.20 ppm in D_2O, compared to δ 4.53 ppm and 4.12 ppm seen in the adenosine analogue[342]). The significantly downfield-shifted H-5′ peaks is particularly characteristic of 5′,8-anhydro-nucleosides: see Chapter 5 for a discussion of this phenomenon. All other sugar peaks are contained between this pair, as they are in the adenosine analogue.[342] Corroboration of this identification comes from comparison of ^{13}C NMR shifts: in both the product formed here, and in the adenosine analogue, the C-5′ peak is deshielded, observed at δ 75.5 ppm and 74.7 ppm respectively.[342]

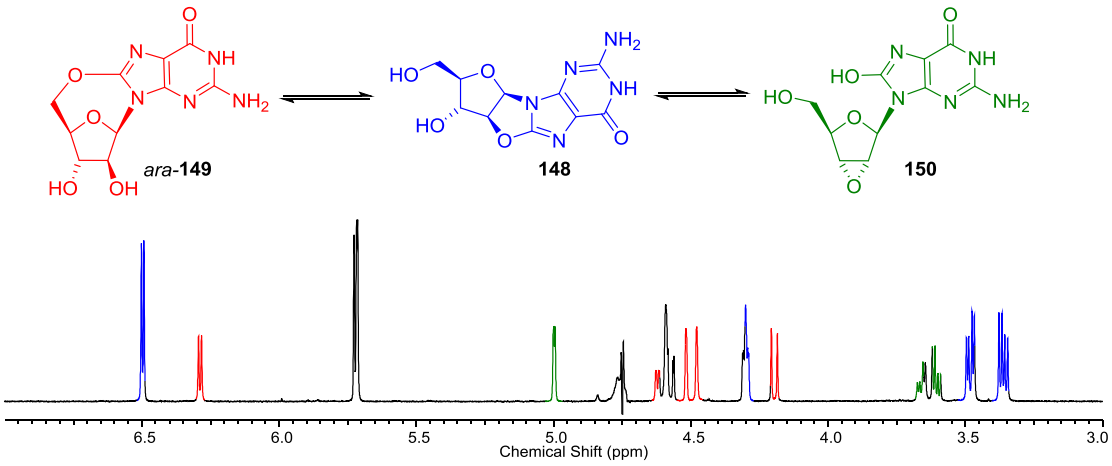


Figure 3.10: *1H NMR spectrum (600 MHz, D_2O, 3.0-9.0 ppm) to show the reaction of 2′,8-anhydroguanosine and boric acid after heating at 50 °C for 3 days (initial pH 6.5) then a 1 day at pH 9.7.*

The product observed to form at higher pH has been assigned as 2′,3′-epoxy-8-oxo-guanosine **150**. It has a characteristic singlet anomeric proton (δ 5.74 ppm in D_2O), confirming that the sugar ring has a strained structure which diminishes the observed coupling to H-2′ by enforcing a constrained, near 90 ° (Karplus) torsional angle; a singlet anomeric proton (δ 6.17 ppm in D_2O) is also seen in the adenosine analogue.[342] Both analogues have the H-2′ peak appearing as a doublet downfield of the water peak with a small coupling constant (δ 4.96 ppm, J = 2.6 Hz for the guanosine derivative and δ 4.56 ppm, J = 2.7 Hz for the known adenosine derivative), and H-5′ peaks, which are relatively close together and upfield shifted. Both adenosine and guanosine derivatives have a distinctively upfield shifted C-2′, appearing at 59.5 ppm (guanosine) and 58.5 ppm (adenosine).[342]

3.4 Conclusion and Outlook

It has been suggested that an alternative to a local high concentration of phosphate on the early Earth is to assume phosphate did not play a part in the formation of the first living system.[182] It is clear that RNA – from its first invention – contained phosphate.[182] Back-extrapolation from modern metabolism also implies that phosphorylated biomolecules were present from early on in life, but the very first organisms may not have used this perfected phosphate-containing system.[182] It is therefore important to seek alternatives to phosphorylated compounds.[346]

Efforts to replace phosphate with borate in the key rearrangement step in an existing prebiotically plausible synthesis of the pyrimidine nucleotides, however, have been unsuccessful. It appears unlikely that borate could have acted in this way to effect stereoinversion at the C′-2 position of the sugar ring with either pyrimidine or related purine derivatives. This is a (partially) disappointing result: borate has been suggested to be a plausible prebiotic reagent, with potential roles in the synthesis and stability of sugars,[147,148,159,160] the purination of ribose,[173] and the formation of nucleobases.[340] Borate-complexation to nucleotides has also been exploited to release phosphate from insoluble minerals[199] and to direct site-selective phosphorylation.[175,199]

Like phosphate, the availability of borate on the early Earth is also disputed.[161] Nevertheless, it is important that all feasible routes towards the biomolecules necessary for life are explored: understanding where reactivity is not observed still helps us understand what plausible chemical scenarios were available on the early Earth.[347] Although borate has been suggested to play a role in the phosphorylation of nucleotides,[175,199] it does not participate in the stereochemical rearrangement of anhydronucleotides. Phosphate, on the other hand, fulfils several essential roles throughout Sutherland's pyrimidine synthesis, not least a key phosphorylation of anhydrocytidine *ara*-**63** – a role that cannot be replaced by borate. A source of phosphate must have been present during the origins of nucleotides.[57]

4. Purines from glycosylation with *O*-acetylated ribose derivatives

4.1 Introduction

Nucleic acids are traditionally retrosynthetically disconnected to deoxy- and ribonucleotides, which in turn are disconnected to a sugar, a nucleobase and phosphate. The synthesis of nucleotides (exemplified by cytidine **29** in ***Figure 4.1***) under prebiotic constraints from these components in significant yield is near-impossible, however, in particular due to the difficulties in forming a glycosidic bond between the sugar and the nucleobase.

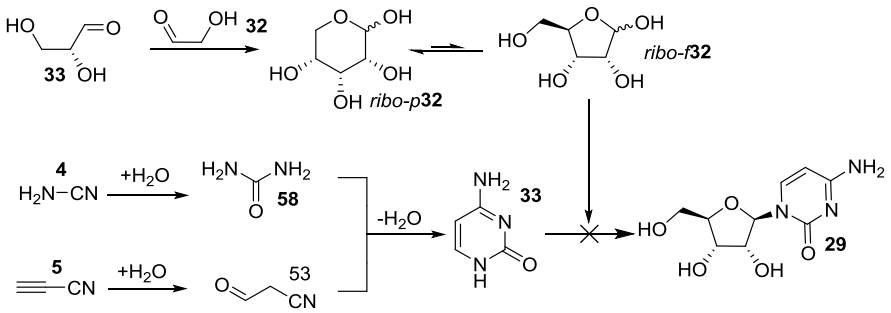


Figure 4.1*: Traditional retrosynthesis of cytidine **29**, via ribose ribo-**30** and cytosine **31**. Glycosylation of **31** by ribo-**30** has not been reported.*

Ribose *ribo*-**30** exists as an equilibrating mixture of different forms in aqueous solution (***Figure 4.2a***); adenine **44** (to use a specific example) also exists as an equilibrating mixture (***Figure 4.2b***).[91] For the desired natural adenosine anomer ß-**55** to form, the anomeric carbon of the α-furanose form of *ribo*-**30** (a minor component in solution) must act as an electrophile: this is possible if the anomeric hydroxyl is protonated under acidic conditions. The *N*-9 atom of adenine must act as a nucleophile, but the adenine tautomer with a lone pair of electrons localised at the *N*-9 atom (**44a**) is only present in low concentrations, whilst the *N*-9 lone pair is delocalised throughout the aromatic system in the major tautomer **44b**. To add further problems, under the acidic conditions that activate α-furanose to reaction with **44**, **44** itself is protonated, giving a form (**44c**) which is not nucleophilic on *N*-9. To add to these problems, the reaction is also thermodynamically disfavoured in water: hydrolysis of the glycosidic bond is favoured over its formation. The prospect for pyrimidine synthesis is even worse: the *N*-1 lone pairs of cytosine **3** and uracil are delocalised around the aromatic ring and into the adjacent carbonyl bonds, whilst the same thermodynamic barrier remains.

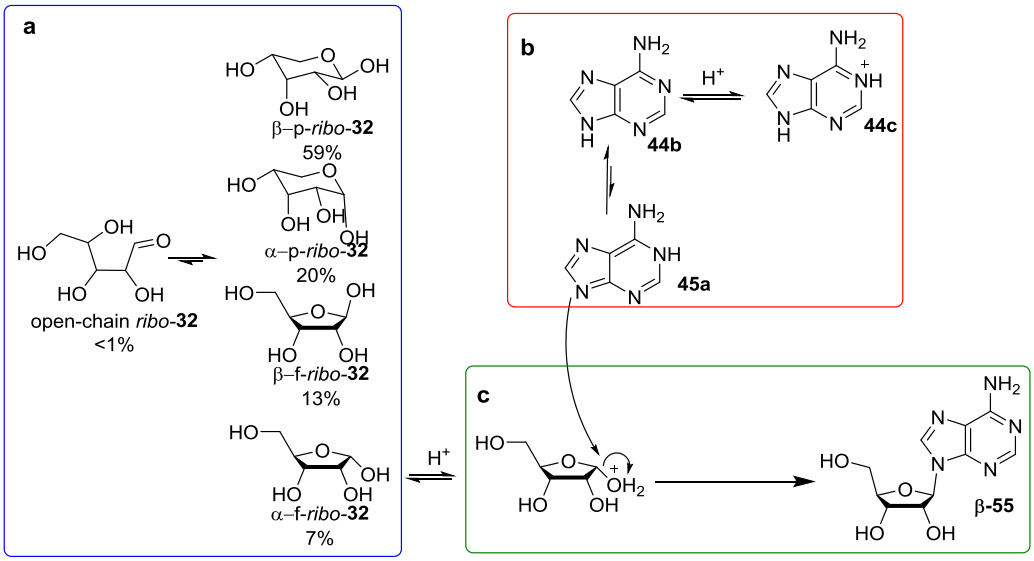


Figure 4.2: *Synthesis of adenosine* ***55*** *under acidic conditions*

a) *ribo-****30*** *anomers present in equilibrium in aqueous solution (blue box)*

b) *Adenine* ***44*** *tautomers present in aqueous solution (red box)*

c) *Synthesis of adenosine ß-****55*** *by glycosylation of adenine with α-furanosyl ribose (green box)*

4.1.1 Nucleosides and nucleotides from glycosylation reactions

Orgel and co-workers attempted numerous glycosylation reactions in aqueous solutions without success. Adenine and ribose in various concentrations failed to produce significant yields of nucleoside products when heated at various temperature (30 to 100 °C) and at a variety of pH values (2 to 11).[172] The addition of condensing agents (including cyanate, cyanamide, urea, and cyanoacetylene), heterogeneous catalysts (clays and minerals) and the presence of UV light also failed to promote condensation.[172] Similar experiments (albeit with a reduced number of variable factors) were also carried out using cytosine, guanine, hypoxathine and uracil, all without success.[172]

Several researchers have investigated glycosylation reactions in dry state: evaporation of the water molecule formed in the condensation of the sugar and nucleobase removes the thermodynamic barrier to this reaction.

Orgel and co-workers attempted a number of reactions in dry state: heating adenine **44** (1 eq.) with ribose *ribo*-**30** (2 eq.) at 100 °C *in vacuo* for 6 h followed by a purification step resulted in the isolation of adenine **44** (74% recovery) and *N*-6-ribosyl adenine **151** (19%); when the adenine/ribose solution is adjusted to pH 4.5 before evaporation and

heating, the yield of **151** is increased to 74% (ie, this reaction appears to be acid catalysed). *N*-6,*N*-9-diribosyl adenine **152** was also isolated: hydrolysis of **7** was then carried out in neutral or alkaline solution, producing ß-**55** in quantitative yield. A 1:2 mixture of α- and ß-adenosine was produced in 0.03% total yield before hydrolysis, and 0.21% after hydrolysis.[172]

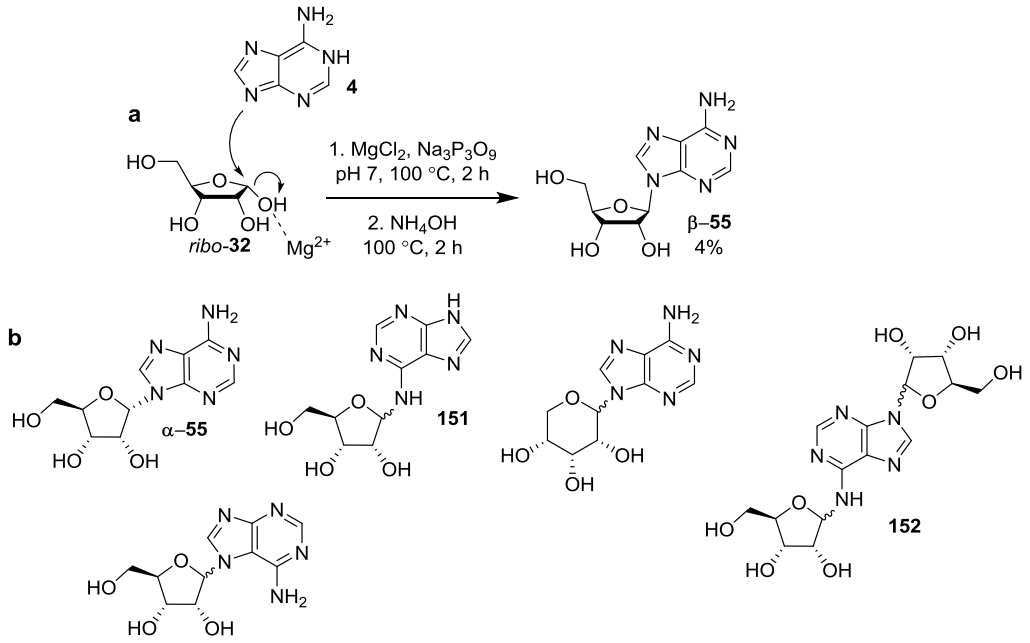


Figure 4.3: *Orgel's synthesis of adenosine* ***55.***

a) *ß-**55** is formed, along with a variety of isomers, by heating ribo-**30** and adenine **44** with sodium trimetaphosphate and magnesium chloride. Hydrolysis of diribosyl adenine **152** under alkaline conditions increases the yield of ß-**55** to 4%.*

b) *A variety of non-natural adenosine analogues are also produced, including **α-55**, N-6-ribosylated adenine **154** and N-9,N-6 diribosylated adenine **155.***

Adenosine **55** has been isolated in higher yields (7.2% after hydrolysis, 1:1 α-**55**:ß-**55**) when adenine **44** (1 eq.) was heated with *ribo*-**30** (6 eq.) in the presence of $MgCl_2$ (5 eq.) and sodium trimetaphosphate (5 eq.), but the *N*-6 ribosylated adenines **151** and *N*-9-ribopyranosyladenine **152** were also formed (***Figure 4.3***). When either $MgCl_2$ or sodium trimetaphosphate were omitted, the yield of adenosine **55** dropped; although $MgCl_2$ could be effectively replaced by $MnCl_2$, its substitution with $CaCl_2$ halved the yield. A number of other salts (including NaCN, $Ca(OH)_2$, $CaCO_3$, $NaSO_4$ and $FeSO_4$) were all found to be ineffective additives in the glycosylation reaction.

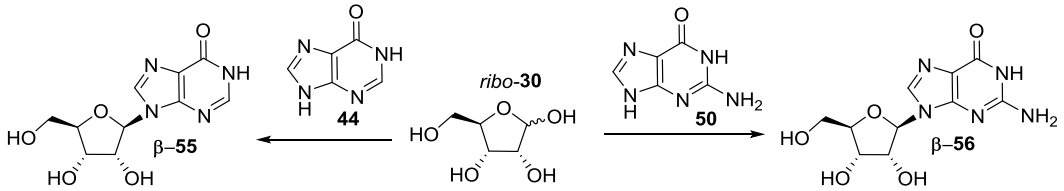


Figure 4.4*: Orgel's synthesis of inosine* ***57*** *and guanosine* ***56*** *by glycosylation of hypoxanthine* ***153*** *or guanine* ***50*** *with ribo-****30****.*

Orgel's method was also applicable to the synthesis of inosine **57** and guanosine **56** (***Figure 4.4***). α-Inosine (4%) and ß-inosine (8%) could be isolated from the reaction of *ribo*-**30** and hypoxanthine **153** in the presence of $MgCl_2$ and sodium trimetaphosphate (without the need for a second hydrolysis step).[172] A slightly higher yield of inosine mixed isomers (13%) could be isolated when *ribo*-**30** (15 eq.) and **153** (1 eq.) were heated in the presence of $MgCl_2$ (2.5 eq.) and $MgSO_4$ (5 eq.); 15% of inosines were isolated when *ribo*-**30** and **153** were heated in the presence of evaporated seawater.[348] ß-guanosine **56** (9%) was the only nucleoside product isolated from the glycosylation and subsequent hydrolysis of guanine **50** and *ribo*-**30** under standard reaction conditions.[172] A recent replication of these reactions, however, failed to achieve the same yields: isolating 3% inosine **57**, and <1% adenosine **55**, in the presence of Mg^{2+}.[349]

A higher yield of ß-**5** (20%) has recently been reported by irradiating *ribo*-**30** and adenine **44** in the presence of formamide and a powdered meteorite with a 170 MeV proton beam.[350] If the powdered meteorite was omitted from the reaction mixture, the yield of the desired product fell to 6%; when adenine **44** and *ribo*-**30** in the dry state were subjected to proton irradiation, the reported yield of ß-**5** was 3%.[350]

Glycosylation under Orgel's conditions only appears to work for purines: attempted glycosylation reactions between cytosine **31**, uracil **51** or thymine and ribose under a variety of conditions failed to produce detectable quantities of pyrimidine nucleosides.[348,349]

As pyrimidine nucleosides cannot be detected under Orgel's glycosidation conditions, various researchers have suggested that the original bases may have differed from the canonical nucleotides known today – and that if these bases are electron-rich, they may react with ribose under glycosylation conditions.

Miller and co-workers have shown that urazole – a five-membered heterocycle synthesised from hydrazine and carbamylurea under prebiotic conditions and possessing a similar hydrogen-bonding ability to uracil – reacts with ribose-5-phosphate *ribo*-**30**-5P in aqueous conditions at mild temperatures (***Figure 4.5***). An insoluble precipitate is formed after heating 'just below reflux' for two days, which is made up of four urazole ribosides **154**: α-furanosyl (13%), ß-furanosyl (22%), α-pyranosyl (11%) and ß-pyranosyl (53%). Equilibration of 100 mM solution of the of four urazole ribosides at room temperature for 86 days resulted in mostly the ß-pyranosyl isomer (78%).[351]

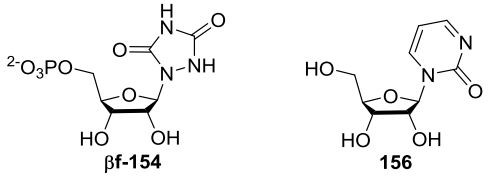


Figure 4.5*: Structures of ß-furanosyl urazole ßf-**154** and ß-furanosyl zebularine **156***

Dry-state heating of 2-pyrimidone **155** with *ribo*-**30** under the conditions described by Orgel (ie 1 eq. nucleobase, 15 eq. ribose, 5 eq. $MgCl_2$ and 10 eq. $MgSO_4$) has been observed to produce ß-furanosyl zebularine **156** in 12% yield, alongside the ß-pyranosyl and α-furanosyl products (***Figure 4.5***).[349] In the absence of magnesium, just 2% **156** was recovered when the solution was adjusted to pH 2.1 before evaporation; none could be detected when the solution was adjusted to pH 6.3.

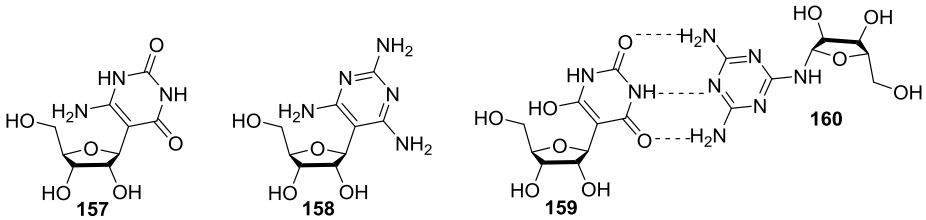


Figure 4.6: *Structures of 6-aminouracil riboside **157**, ß-furanosyl-2,4,6-triaminopyrimidine **158**, and barbituric acid **159** Watson-Crick base-paired with melamine **160**.*

Further nucleoside analogues can be made if researchers do not restrict themselves to making natural *N*-glycosides (***Figure 4.6***). Under Orgel's glycosylation conditions, Benner and colleagues have made a *C*-glycoside **157** from 6-aminouracil and an excess of *ribo*-**30** (no yields are specified);[352] and Hud and colleagues have made a ß-ribofuranosyl nucleoside **158** containing a *C*-glycosidic bond in 60-90% yield when *ribo*-**30** and the pyrimidine derivative are heated in the absence of any salts.[353] Hud and co-workers have also shown that two (non-natural) nucleotides spontaneously form in

aqueous solution from ribose-5-phosphate *ribo*-**30**-5P and the respective heterocycle: neither nucleotide contains a natural *N*-glycosidic bond (barbituric acid **159** contains a *C*-glycosidic bond, and the glycosylation of melamine **160** occurs at an exocyclic amine), although the two are capable of Watson-Crick base pairing (which, in fact, canonical nucleotide monomers are incapable of doing).[354]

4.1.2 UV irradiation of acetylated nucleotides

It is known that cytidine **29** and uridine **60** form a complex mixture of products after UV irradiation.[253,254] Experiments nearly 50 years ago showed that α-**29** is converted to the natural ß-isomer in just 4% yield by photoanomerisation, and that ß-cytidine-5′-phosphate is formed in 6% yield from the α-isomer.[253] It was not clear if the low yields in these photoanomerisations were due to low conversions, or to the formation of by-products, so this photoanomerisation has recently been extensively investigated.[254] This revealed that the low yield of the desired ß-isomers are as a result of completing nucleobase destruction and nucleobase loss: cytosine **31**, uracil **51** (albeit in trace amounts) and ribose oxazolidinone *ribo*-**64** are the only products detectable by ^{1}H NMR when α-**29** is subjected to UV irradiation for 72 h. The same degradation products, as well as oxazolidinone-5′-phosphate, are detected when α-cytidine-5′-phosphate is irradiated for 15 h. The sugar products from the nucleobase loss resulting from the irradiation of nucleosides could not be identified.

The authors presented a possible mechanism for the photo-destruction of the nucleosides (***Figure 4.7***). The nucleobase moiety of α-**29** initially forms a Dewar-pyrimidine structure **161** upon irradiation. The enamine functionality of the Dewar structure can undergo C-protonation to give a bicyclic iminium **162**. Intramolecular attack of the 2′-OH on the strained (and therefore reactive) carbonyl group results in oxazolidinone *ribo*-**64** formation. Alternatively, electrocyclic ring-opening, promoted by the charge and strain in the bicyclic Dewar structure, generates a monocyclic iminium ion **163** which can be quenched with the hydroxide counter ion to provide a photohydrate **164**. The cytidine-photohydrate **164** undergoes competing elimination of water (re-forming α-cytidine **29**) or hydrolysis, giving the uridine-hydrate **24**, from which the elimination of water provides uridine α-**60**. Irradiation of α-**60** would then be expected to form a similar Dewar stucture. The high reactivity of the carbonyl and partial positive charge on the *N*-1 atom in the iminium ion intermediates (of both

cytidine and uridine) means that the nucleobase moiety is a good leaving group, releasing cytosine **31** and uracil **51** respectively.

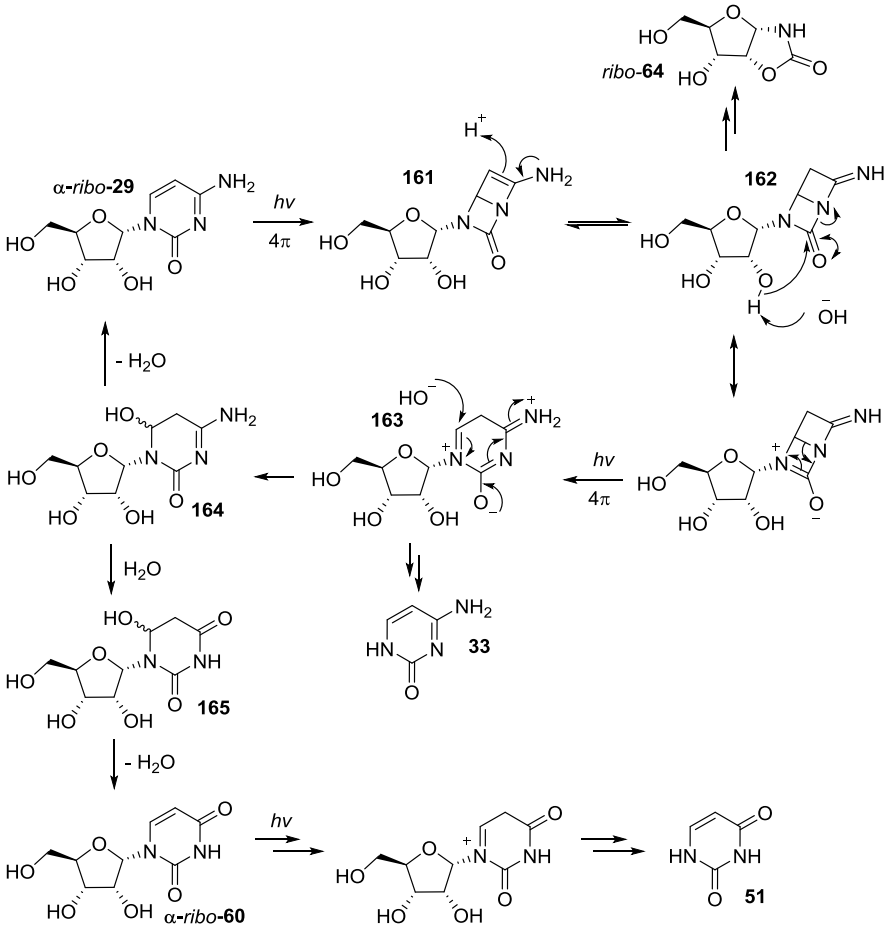


Figure 4.7: *Proposed mechanism for the photo-destruction of α-cytidine.α-Cytidine (α-**29**) forms a Dewar pyrimidine structure **162**, which can undergo intramolecular attack to release oxazolidinone ribo-**64**, or ring-opening to provide photohydrate **164. 164** can undergo elimination of water (reforming α-**29**) or hydrolysis, releasing uridine-hydrate **24**.*

As significant nucleoside loss occurs through oxazolidinone **22** formation, *via* intramolecular attack of the 2′-OH group on the Dewar pyrimidine, it appears obvious that this deleterious reaction pathway cannot take place when the 2′-OH group is absent or blocked. Accordingly, Powner and colleagues investigated the irradiation of α-cytidine-2′-phosphate. As expected, a higher proportion of nucleotide products, and less oxazolidinone, were seen: preventing by-product formation allowed a longer irradiation time, with increased amounts of photoanomerisation and photohydrolysis observed. Some nucleobase loss was also observed, but, unlike the irradiation of nucleosides, ribose-2-phosphate could be identified in the reaction mixture.

Acetylated pyrimidines behave differently to the nucleosides and nucleotides already studied: acetylation of α-cytidine-5′-phosphate α-**29**-5'P protects the *cis* diol, preventing oxazolidinone formation and C-2′ epimerisation and enhancing photoanomerisation to the ß-pyrimidine.[355] Prolonged UV irradiation of 2′,3′-*O*-diacetylated α-cytidine-5′-phosphate **166**, however, results in base loss releasing the diacetylated sugar derivative **167** (***Figure 4.9***). Protection of the 2′-OH group means that oxazolidinone formation is prevented; the acetyl protecting groups also presumably help prevent photodestruction of the free sugar moiety.

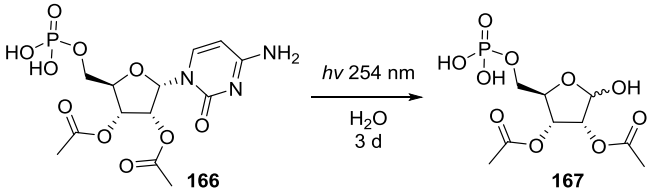


Figure 4.8: *De-pyrimidation of acetylated α-cytidine* ***166*** *after prolonged UV irradiation.*

Chemoselective acetylation in water under prebiotic conditions has been studied as a way to facilitate nucleotide ligation and improve natural 3′,5′-linkage homogeneity.[139] 2′-OH acetylation preferentially occurs when a mixture of nucleotide 2′- and 3′-phosphates are treated with sodium thioacetate **27** and cyanoacetylene **5** (although a precise explanation for this selectivity is unclear). When acetylated oligomers are activated and ligated, the 2′-protecting group prevents the rapid formation of a 2′,3′-cyclic phosphate terminus, promoting instead 3′,5′-ligation; subsequent ammonolysis removes the acetyl groups to generate native RNA without significant backbone hydrolysis (***Figure 4.9***).

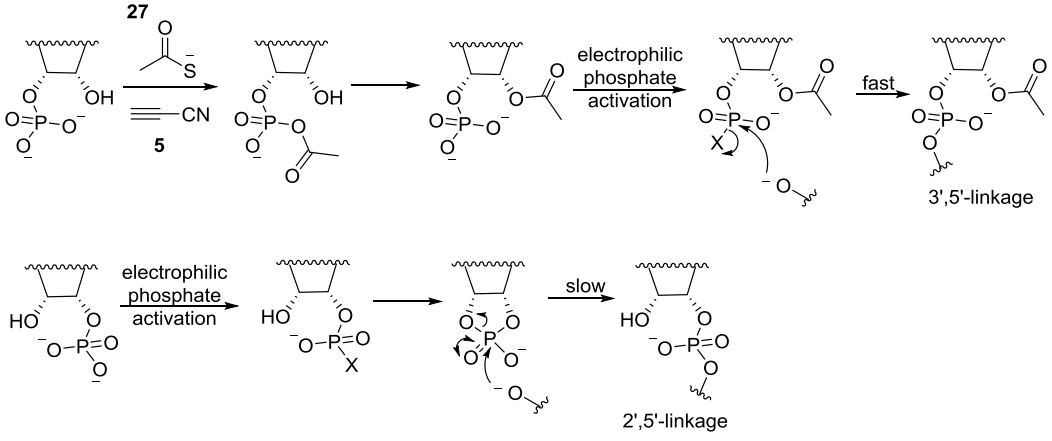


Figure 4.9: *Chemoselective acetylation of 3′-phosphate terminal oligomers in the presence of corresponding 2′-terminal oligomers. The free 2′-OHs are selectively acetylated in the presence of* ***27*** *and* ***5****, protecting them against cyclisation to 2′,3′-cyclic phosphate termini upon electrophilic phosphate activation.*

4.2 Project aims

Prebiotic synthesis of natural pyrimidine nucleotides relies on an irradiation step to partially convert ß-cytidine-2′,3′-cyclic phosphate to ß-uridine-2′,3′-cyclic phosphate and also clean up the reaction mixture by destroying the other nucleosides and nucleotides present (including α-isomers). We suggest that these unwanted α-isomers could be recycled instead of destroyed. Acetylation and irradiation of these unwanted α-by-products would result in the release of an acetylated ribose derivative, **167**. A glycosidation reaction could take place between **167** and a nucleobase base to selectively give *O*-acetylated ß-nucleotides: the acetyl protecting groups of **167** are expected to block attack from the α-face of the sugar, and also help prevent degradation of the sugar under the harsh reaction conditions. This could provide a parsimonious route to natural ß-purine nucleotides by recycling unwanted α-pyrimidines (***Figure 4.10***). One of the criticisms of Powner's work is that the cyanamide and cyanoacetylene required are highly reactive (and therefore unstable) species; purine nucleobases are, by contrast, more stable and could be prebiotically available.[175] The exact mechanism of the glycosylation reaction is unclear; however, whether purination occurs by S_N2 displacement of the anomeric hydroxyl group (upper intermediate) or by attack on an oxycarbenium ion (lower intermediate), the 2′-acetyl groups should provide effective protection of the α-face of the sugar (similar to the protection of the α-face of the sugar by 2′-esters in the Hilbert-Johnson glycosylation[356,357]).

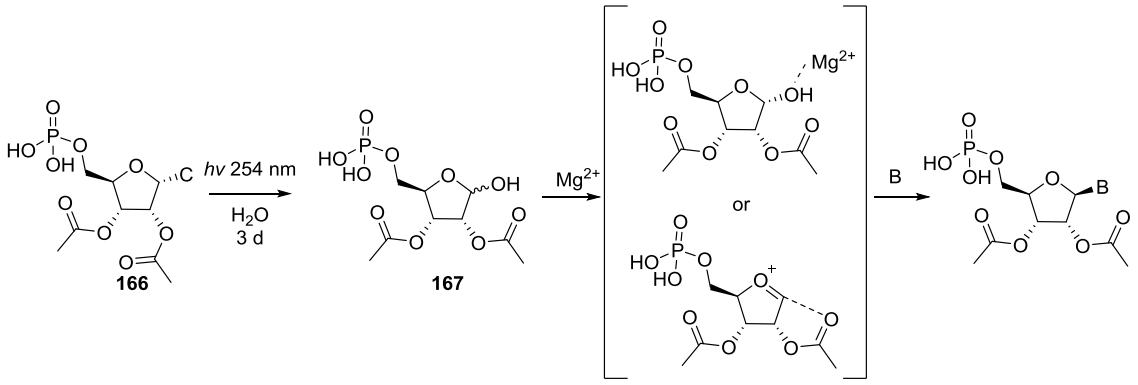


Figure 4.10: *Proposed parsimonious prebiotic route to natural ß-purine nucleotides by recycling unwanted α-pyrimidines. Preliminary investigation has revealed a prebiotic route from **166** to **167**. The acetyl protecting groups block purine attack on the lower face of the sugar, directing the formation of a ß-glycosidic bond. (C = cytosine **31**, B = base)*

4.3 Results and Discussion

4.3.1 Synthesis of protected sugars

The necessary acetylated ribose derivatives for this study were first synthesised using a variety of methods. The standard procedure (acetic anhydride in pyridine) was used for the per-acetylation of *ribo*-**30** with good results, giving **169** in 71% yield (***Figure 4.11a***). An acetylation protocol developed and optimised by Fernandez-Garcia[358] for the preparative synthesis of selective *O*-acetylated nucleotides using *N*-acetylimidazole **168** was adapted for the acetylation of ribose-5-phosphate *ribo*-**30**-5P.Two plausible mechanisms can be envisaged for this acetylation: the 3′-hydroxyl group could be protected after acetyl transfer from a mixed anhydride (as has been proposed for other aqueous acetylations[139]), or by direct hydroxyl acetylation (***Figure 4.11b***).

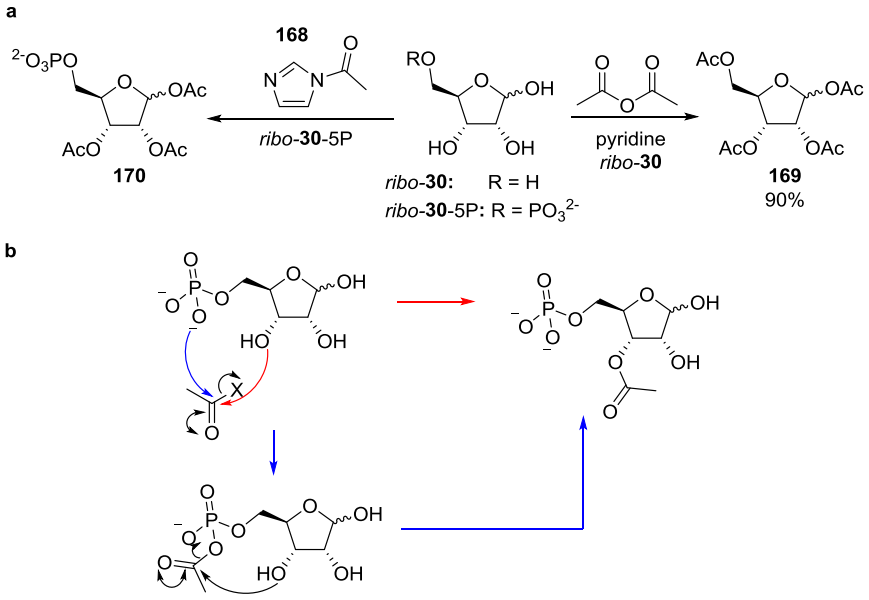


Figure 4.11: *Acetylation of ribose ribo-**30** and ribose-5-phosphate ribo-**30**-5P*

a: *The standard method (acetic anhydride in pyridine) was used for the per-acetylation of ribo-**30** with good results; an acetylation protocol for the preparative synthesis of acetylated nucleotides was adapted for the acetylation of ribo-**30**-5P.*

b: *Potential pathways for acetylation of ribo-**30**-5P with N-acetyl imidazole **168:** intramolecular acyl transfer from mixed anhydride (blue arrows); direct acetylation of hydroxyl (red arrows).*

High conversion to acetylated ribose-5-phosphate products (92% by ^{1}H NMR relative integration) occurred within 2 h, but 1,2,3-tris-acetyl-ribose-5-phosphate **170** proved difficult to purify. Lyophilisation of the reaction mixture appeared to cause partial de-acetylation. 1,2,3-tris-acetyl-ribose-5-phosphate **170** was eventually obtained as a mixture of both α- and ß-anomers in just 20% isolated yield after flash column chromatography and reverse-phase C18 flash column chromatography. Other fractions

also contained the tris-acetylated product **170** as a mixture with the mono- and bis-acetylated sugars.

Deprotections of the C-1 position of these acetylated sugars were also investigated. A number of literature routes to 2,3,5-tri-*O*-acetyl ribose **171** are available: Carell *et al.* for example, generate this partially protected sugar in 30% over three steps from ribose *via* benzylation at C-1,[359] acetylation of the remaining hydroxyl groups and reductive cleavage of the benzyl group.[360] As per-acetylated ribose had already been synthesised as part of this study, however, it appeared prudent to explore 1-deprotection strategies. There are several ways to effect this transformation in literature: for example, with lipase (63% yield)[361] and tributyltin methoxide (80%).[362] A 1-deprotection protocol using hydrazine and DMF (both readily available) and previously employed by other Powner group members was explored. Unfortunately, attempts to use this procedure in this project proved ineffective: the desired product **171** (yield <20%) could not be isolated from a mixture of by-products, although starting material **170** (23%) was recovered from the reaction.

Previous work in the Powner group had inadvertently resulted in the production of 2,3,5-tri-*O*-acetyl ribose **171** during an attempted synthesis of the 1,2-alkene derivative **172** (***Figure 4.12***). We thought this reaction could be optimised to provide our desired product, **171.**

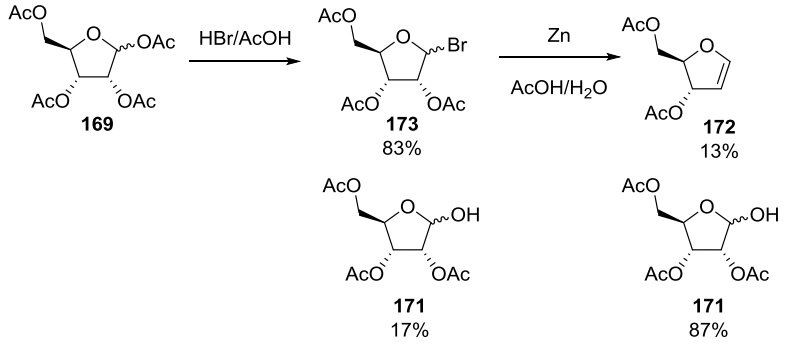


Figure 4.12: *Inadvertent production of 2,3,5-tri-O-acetyl ribose **171** achieved during a synthesis of **172** attempted by another member of the Powner group.*

Using an allusion in a paper by Le Merrer and co-workers[363] to the selective hydrolysis of the anomeric position of peracetylated ribose by successive bromination and treatment with sodium hydrogen carbonate as a starting point, we developed a protocol for the selective 1-deprotection of the per-acetylated ribose derivative **7**. This one-pot

two-step procedure afforded 2,3,5-tri-*O*-acetyl ribose **171** in 30% yield after purification (***Figure 4.13***).

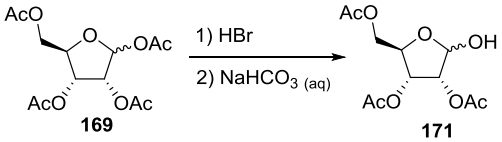


Figure 4.13: *One-pot two-step synthesis of 2,3,5-tri-O-acetyl ribose **171** from **169**.*

These *O*-acetylated ribose derivatives were used to investigate the requirements for and role of the sugar in Orgel's prebiotic glycosidations.

4.3.2 Glycosylation reactions

It was decided that inosine **57** should be the focus of investigations: although Orgel produced guanosine **9** in higher yields (9%, versus 8% achieved for **57**), the insolubility of guanine **50** caused the researchers some problems when setting up and analysing the reaction.[172] Hud has since replicated Orgel's purine glycosylations using adenine **44** and hypoxanthine **153** and achieved much lower yields, but the yield of inosine **57** (3%) is nevertheless higher than that of adenosine **55** (<1%).[349] Furthermore, **153** has only one site at which glycosylation can take place, so the reaction products do not require a hydrolytic work-up, unlike the products of the reaction of **44** and *ribo*-***30***.[172]

Orgel's synthesis of inosine **57** was replicated with good results: after heating for 4 h, approximately 7% conversion from hypoxanthine **153** to inosine was estimated by relative integration of these species in ^{1}H NMR. The presence of inosine ß-**57** was confirmed by spiking with a commercial sample (***Figure 4.14***).

A programme of investigation of the reaction of hypoxanthine **153** with acetylated ribose derivatives was undertaken. No nucleoside products were observed by ^{1}H NMR spectroscopy following dry-state heating of 1,2,3-tris-acetyl-ribose-5-phosphate **170** with **153** under Orgel's reported glycosylation conditions, or in the reaction of 2,3,5-tris-acetyl-ribose **171** with **153** under glycosylation conditions.

The suitability of iron as a Lewis acid catalyst in this reaction was also tested by carrying out glycosylation reactions between 2,3,5-tris-acetyl-ribose **171** and hypoxanthine **153** in the presence of $FeCl_2$ and $FeCl_3$, but no nucleoside was seen by ^{1}H NMR in either of these reactions. This corroborates the findings of Orgel, who

reported that ferrous chlorides did not promote the glyosidic formation of adenosine **55**.[172]

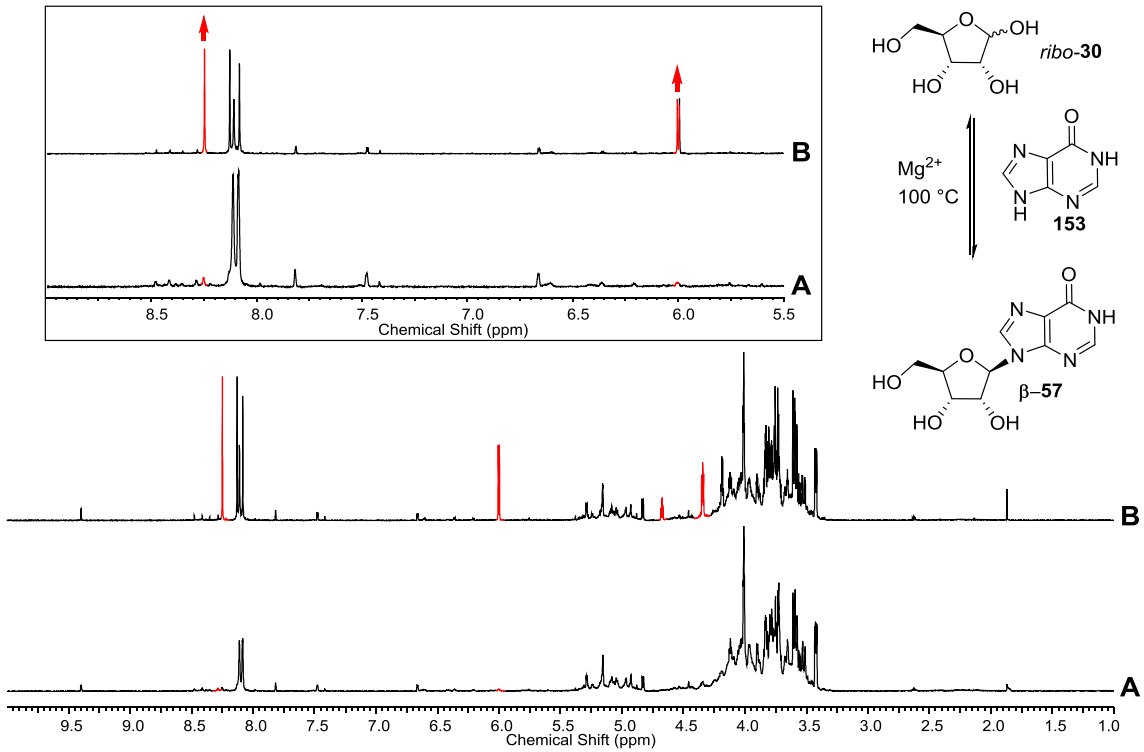


***Figure 4.14**: [1]H NMR spectra (600 MHz, D_2O, 1.0-10.0 ppm) showing synthesis of inosine ß-**57** (highlighted in red) under glycosylation conditions. **A**) Ribo-**30** (20 eq.), hypoxanthine **153** (1 eq., then an additional 1 eq. after 2 h heating), $MgSO_4$ (10.5 eq.) and $MgCl_2$ (5 eq.) after heating at 100 °C for 4 h; and **B**) spiked with ß-**57**. A number of peaks in the same region are likely to be analogues of inosine (eg α-furanosyl and pyranosyl isomers).*

We postulated that the lack of success in these glycosylation reactions may in fact be due to the presence of protecting groups on the ribose derivatives. Neither 1,2,3-tris-acetyl-ribose-5-phosphate **170** nor 2,3,5-tris-acetyl-ribose **171** has a free 2′-OH moiety – perhaps this is required to correctly coordinate the Mg^{2+} ion to allow the glycosylation reaction to occur. Hud's computational modelling of the reaction of *ribo*-**30** and 2-pyrimidone **155** leads him to state that an interaction is observed between Mg^{2+} and 2-OH during glycosidation.[327] Orgel, however, describes preliminary results indicating that ß-deoxyadenosine **174** (<2%) and ß-deoxyinosine **175** (>5%) are formed when deoxyribose **176** and either adenine **44** or hypoxanthine **153** are heated in sea water.[348] Our attempts to reproduce the synthesis of deoxynucleosides under glycosylation conditions were unsuccessful: after heating **176** and **44** in the presence of Mg^{2+} at 100 °C for 7 h, a number of new peaks were seen in the aromatic region of the

^{1}H NMR spectrum as well as a new peak in the H-1′ region, but spiking with 2′-deoxyadenosine **174** indicated this had not been formed.

The glycosylation reaction using ribose-5-phosphate *ribo*-**30**-5P, which has a free 2-OH moiety, was also explored. *ribo*-**30**-5P and hypoxanthine **153** did not form inosine-5-phosphate ß-**57**-5'P under glycosylation conditions, even after heating at 100 °C for 17 h. Although a number of new peaks were seen in the aromatic region of the ^{1}H NMR, as well as a new peak in the H-1′ region, when adenine **44** and *ribo*-**30**-5P were heated under glycosylation conditions, spiking with adenosine-5′-phosphate ß-**55**-5'P indicated this species had not been formed. Orgel reported that phosphorylated sugars do not undergo glycosylation. Although Orgel recovered phosphorylated ribose from the attempted glycosylation of *ribo*-**30** and adenine **44** in the presence of $MgCl_2$ and sodium trimetaphosphate, no phosphorylated derivatives of adenosine could be isolated. Orgel heated the phosphate-containing products he recovered from this reaction with adenine **44** under glycosylation conditions, but adenosine (and, presumably, its phosphorylated derivatives) was not formed. α- and ß-Ribofuranose-1-phosphate were also reportedly unreactive under these conditions.[172]

Although arabinose *ara*-**30** has a free 2-OH moiety, it is on the opposite face of the sugar ring to the α-anomeric hydroxyl, and therefore the two hydroxyl groups cannot co-ordinate Mg^{2+} simultaneously (when the anomeric hydroxyl is in the ß-orientation an α-nucleotide will be formed through S_N2 displacement). The glycosylation reaction between arabinose *ara*-**30** and adenine **44** was attempted, but arabinose adenosine *ara*-**55** was not detected by ^{1}H NMR analysis among the products of the reaction.

It was noted that the pH of the residue of these reactions was lower when resuspended in water than before evaporation at the start of the reaction. It is unclear if this drop in pH is a result of the sugar decomposition seen over the course of the reaction, or if the drop in pH causes the decomposition of the sugar. However, the decomposition of ribose-5-phosphate *ribo*-**30**-5P under glycosylation conditions appears to be less severe than ribose decomposition: it is possible that the phosphate moiety buffers the reaction pH. These glycosylation reactions were therefore repeated in the presence of inorganic phosphate, but no nucleoside could be detected by ^{1}H NMR spectroscopy in any of these reactions, indicating the phosphate inhibits glycosylation by preventing acidification or metal ion binding inhibiting either specific acid or Lewis acid catalysis.

4.3.3 Decomposition of nucleosides and nucleotides

The most easily formed nucleosides are also expected to be those that have the most labile bonds under the same reaction conditions: Hud and colleagues observed a positive correlation between nucleoside degradation and nucleoside synthesis yields (significantly, decomposition of uridine **60** was not observed).[349] Orgel, meanwhile, discovered that when hypoxanthine **153**, ribose *ribo*-**30**, and ^{14}C-labelled inosine **57** were heated with evaporated seawater, only 43% of ^{14}C-**57** was recovered, whilst 30% was hydrolysed to **153** and 27% was converted to two unidentified products.[348] We therefore decided to heat **57** and its acetylated derivatives under glycosylation conditions to explore their degradation behaviour, which could provide an insight into the problems seen in synthesising them.

Inosine **57**, when heated under glycosylation conditions in the presence of Mg^{2+} was observed to partially degrade over the course of 26 h: the major decomposition product was characterised by the presence of two singlet peaks in the aromatic region. The reaction was repeated in the presence of phosphate buffer. Again, **57** partially degraded to give a single major decomposition product characterised by singlet peaks at δ 8.09 and δ 8.06 (degradation was estimated at 19% conversion by relative integration of starting material and product peaks). Spiking with hypoxanthine **153** confirmed the presence of this nucleobase (***Figure 4.15***). The decomposition of inosine **57** in the presence of $FeCl_2$ and $FeCl_3$ was also investigated: in each case, **57** was the only species observed by ^{1}H NMR spectroscopy after heating at 100 °C for 2 h.

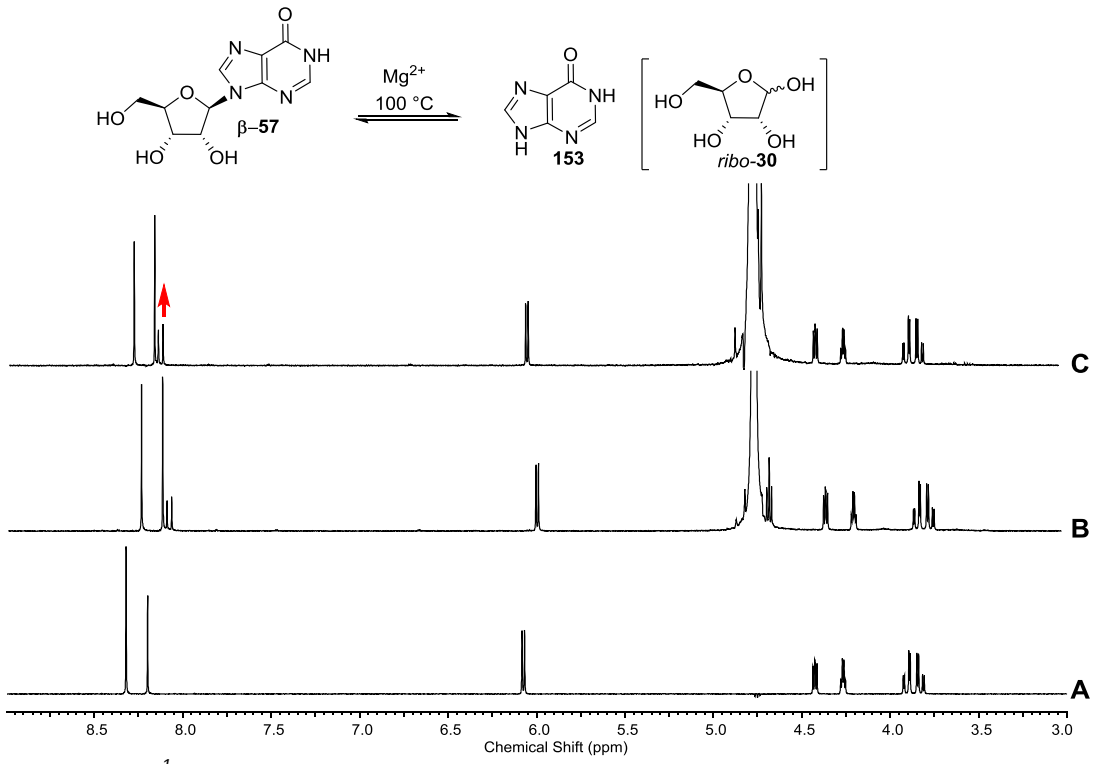


***Figure 4.15**: ^{1}H NMR spectra (600 MHz, D_2O, 3.0-9.0 ppm) showing decomposition of inosine **57** under glycosylation conditions in the presence of phosphate buffer. **A**) inosine **57** (3.0 eq.); **B**) after heating with $MgSO_4$ (2 eq.), $MgCl_2$ (1 eq.) and Pi (1.17 eq) at 100 °C for 5 h; and **C**) spiked with hypoxanthine **153**.*

The decomposition of 2′,3′,5′-acetylated inosine **177** was next investigated (***Figure 4.16***). **177** was not observed to degrade in the presence of Mg^{2+}, even after heating at 100 °C for 30 h; neither was **177** observed to degrade in the presence of stoichiometric or excess inorganic phosphate over 8 h at 100 °C. When heated for a further 4 h (2.2 eq Pi), ^{1}H NMR analysis indicated that two nucleoside species were present in a 1:1 ratio, but only five acetyl groups were present, indicating the greater lability of the acetyl protecting group than the glycosidic bond. **177** was the only species observed by ^{1}H NMR spectrosopy after heating at 100 °C in the presence of Fe^{2+} and Fe^{3+} for 12 h. These results suggest that a glycosylation reaction between 2′,3′,5′-acetylated ribose **171** and hypoxanthine **153** under these conditions is unlikely.

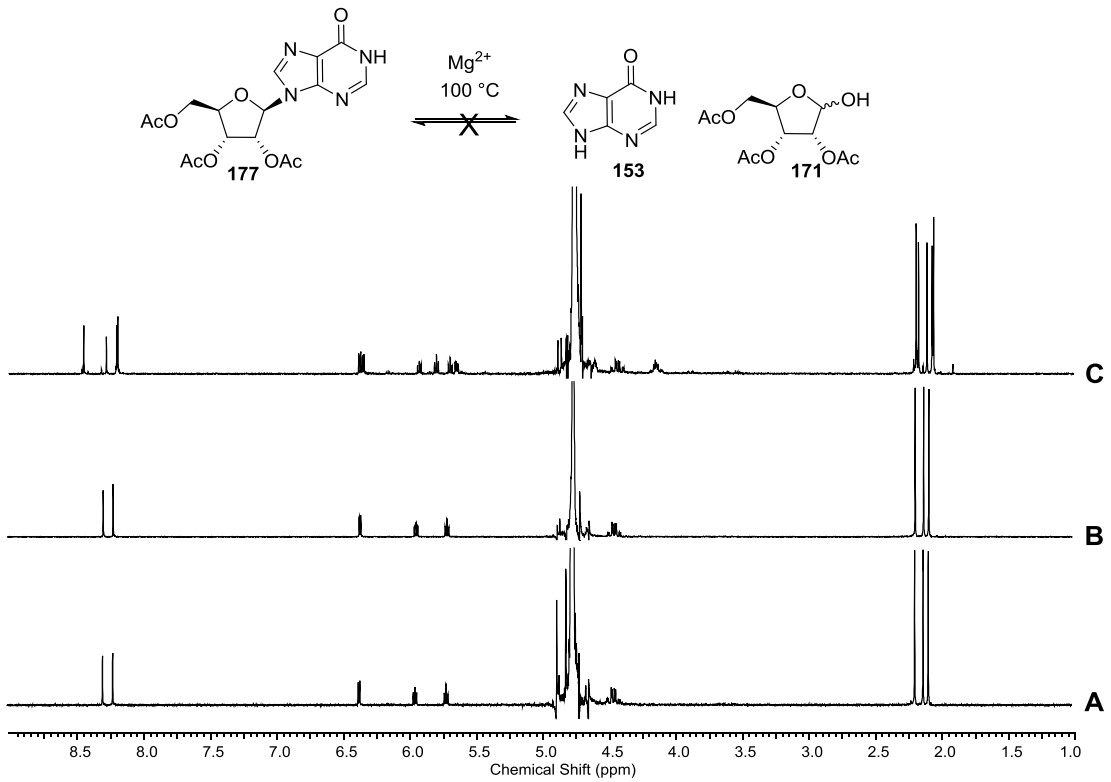


***Figure 4.16**: ^{1}H NMR spectra (600 MHz, D_2O, 1.0-9.0 ppm) showing decomposition of per-acetylated inosine **45** under glycosylation conditions. **45** (3 eq.), $MgSO_4$ (2 eq.) and $MgCl_2$ (1 eq.) after heating at 100 °C for: **A**) 30 h; **B**) 30 h in the presence of inorganic phosphate (4 eq.); and **C**) 12 h in the presence of inorganic phosphate (6.4 eq.).*

No decomposition was observed when 2′,3′-acetylated inosine-5′-monophosphate **178** was heated under glycosylation conditions in the presence of Mg^{2+}, even after 30 h at 100 °C. In the presence of increasing amounts of inorganic phosphate, however, **178** degraded to give a complex mixture of products: six other species, including inosine-5′-monophosphate ß-**57**-5'P, were formed in the presence of 1 eq. inorganic phosphate; inosine **57** was also observed in the presence of 2 eq. inorganic phosphate (***Figure 4.17**).* The other products observed were not identified, but are likely to include partially acetylated derivatives of the identified species. The same species were also observed, but formed much more quickly, when 2′,3′-acetylated inosine-5′-monophosphate **178** was heated in the presence of inorganic phosphate without divalent cations. Again, these observations indicate that the glycosidic bond of 2′,3′-acetylated inosine-5′-monophosphate **178** is relatively stable, and will not be easily formed under glycosidation conditions.

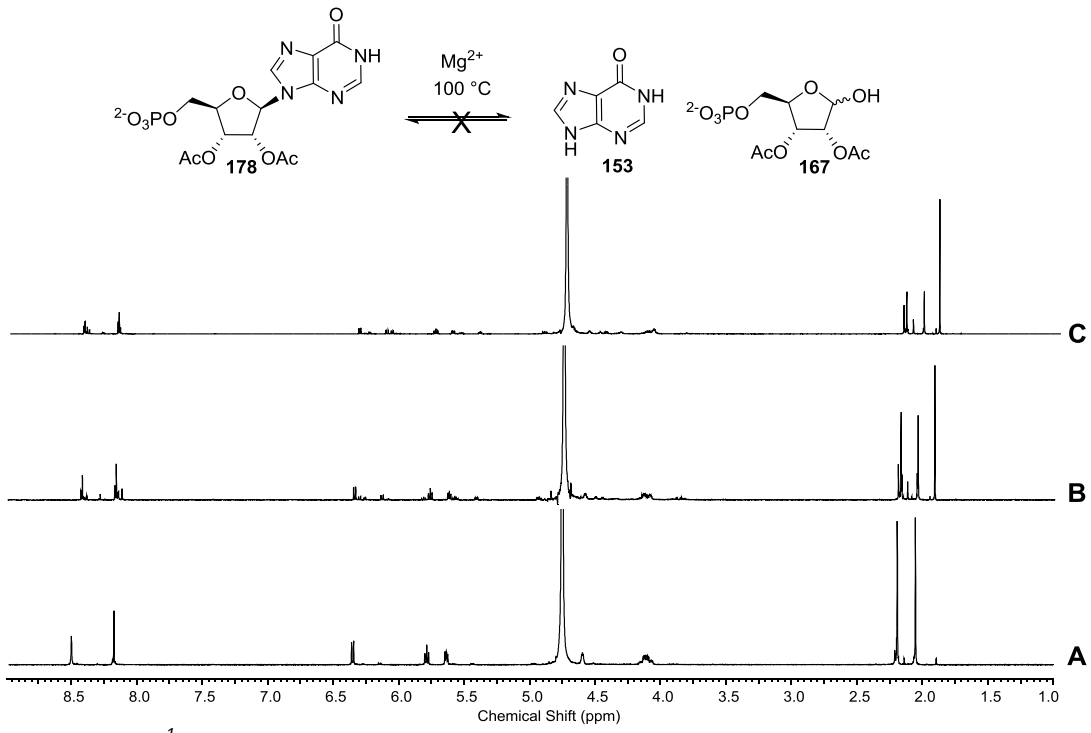


***Figure 4.17**: [1]H NMR spectra (600 MHz, D_2O, 1.0-9.0 ppm) showing decomposition of 2',3'-acetylated inosine-5'-monophosphate **178** under glycosylation conditions. **178** (3 eq.), $MgSO_4$ (2 eq.) and $MgCl_2$ (1 eq.) after heating at 100 °C for: **A**) 30 h; **B**) 12 h in the presence of inorganic phosphate (0.9 eq.); and **C**) 2 h in the presence of inorganic phosphate (4 eq.).*

As one of the decomposition products of 2′,3′-acetylated inosine-5′-monophosphate **178** was inosine-5′-monophosphate ß-**57**-5'P, the decomposition of this species of this species under glycosylation conditions was also investigated. In the presence of Mg^{2+}, a small amount of inosine **57** (5% conversion estimated from relative integration of ß-**57**-5'P) was formed after heating at 100 °C for 50 h; in the presence of Mg^{2+} and inorganic phosphate (1.1 eq.), three species were observed by ^{1}H NMR spectroscopy after heating at 100 °C for 5 h (67% ß-**57**-5'P, 19% inosine ß-**57** and 14% hypoxanthine **153**).

The decomposition of adenosine ß-**55** and its derivatives were also explored. ß-**55** was stable under glycosylation conditions. (This reflects the relative ease of inosine **57** synthesis compared to **55** synthesis reported by Hud and Orgel.[172,348,349]) Adenosine-5′-monophosphate ß-**55**-5'P partially decomposed over 7 h giving 6% conversion to adenosine ß-**55**; several new aromatic peaks were seen in the product ^{1}H NMR spectrum when 2′-deoxyadenosine **174** was heated at 100 °C for 7 h in the presence of

Mg^{2+}. A number of new aromatic peaks were also seen in the product ^{1}H NMR spectrum when arabinose adenosine **44** was heated under glycosylation conditions.

4.3.4 Glycosylation of Hilbert-Johnson bases

The acetylated sugar derivatives observed as a by-product of acetylated nucleotide irradiation used in the preceding reactions are remarkably similar to those commonly used in the well-established Hilbert-Johnson glycosylation reaction. Hilbert and Johnson discovered that nucleoside analogues could be made from the reaction of a 1-bromo-acetylated sugar and a pyrimidine.[356,357] Acetyl protection at the C-2 position of ribosyl sugars protects the bottom face from attack (by forming a cyclic cationic intermediate), directing glycosylation to the ß-face, and consequently yielding ß-nucleosides (***Figure 4.18***).

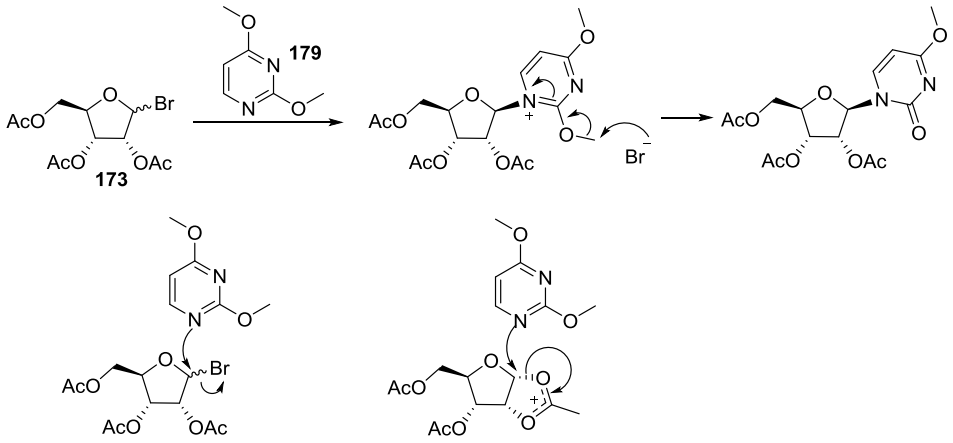


Figure 4.18*: The Hilbert-Johnson synthesis of pyrimidine nucleotide analogues using 2,4-methoxy pyrimidine* ***179****. Two plausible reaction intermediates are shown: direct displacement of a bromide ion, and via formation of a cyclic cation.*

Accordingly, we wondered whether the Hilbert-Johnson pyrimidine nucleobase analogues would react under our glycosylation conditions with ribose *ribo*-**30**. The glycosylation reaction of 4-amino-2-methoxypyrimidine **180** and *ribo*-**30** was attempted with a variety of initial pH values (pH 4.5–7) but all without success: instead it was found that **180** decomposed under reaction conditions to give cytosine **31** (***Figure 4.19***).

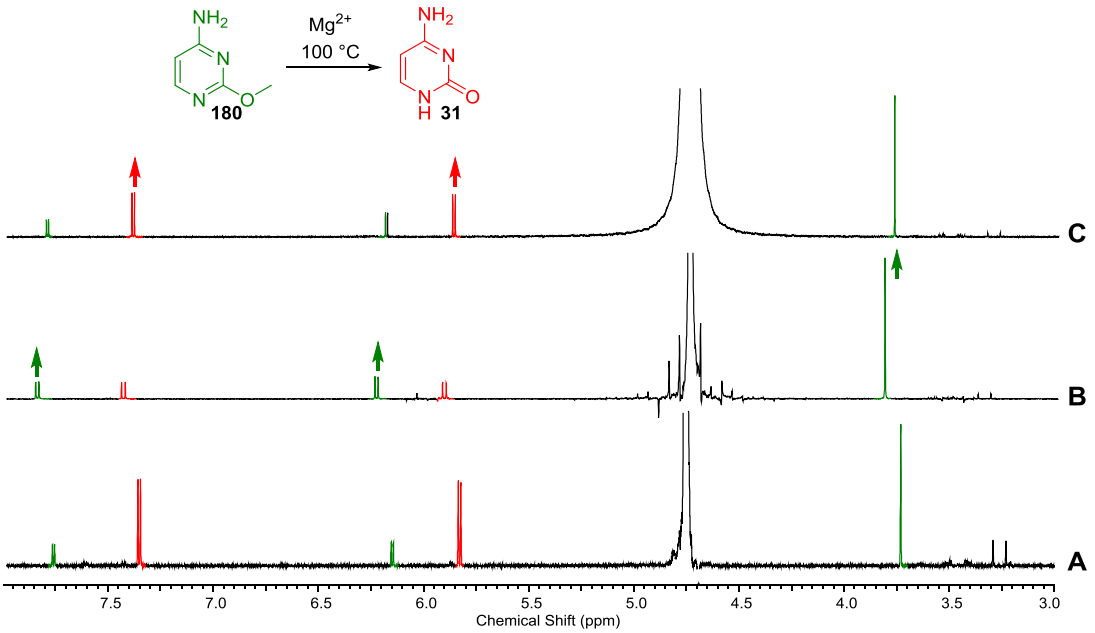


***Figure 4.19**: ^{1}H NMR spectra (600 MHz, D_2O, 3.0-8.0 ppm) showing decomposition of 4-amino-2-methoxypyrimidine **180** under glycosylation conditions. **A**) **180** (1 eq.), $MgSO_4$ (10 eq.) and $MgCl_2$ (5 eq.) heated at 100 °C for 2 h. 79% conversion to cytosine **31** is observed. **B**) spiked with **180**; **C**) spiked with **31**.*

4-Amino-2-(methylthio)pyrimidine **181** appeared, by contrast, to be stable under the conditions employed. In the presence of ribose no nucleoside product was observed at higher pH values (pH 5.5-7.0), but at lower pH values (pH 4.5-5.0) a new unidentified product was observed, characterised by the presence of a doublet at δ 6.63 ppm (J = 7.3 Hz, 1H) and a broad singlet at 7.98 ppm (1H) in the ^{1}H NMR spectrum. The proportion of this new product did not increase over time.

4.4 Conclusions and outlook

From the experiments carried out with a view to investigating a possible formation of a glycosidic bond between acetylated ribose derivatives (observed by-products of the UV irradiation of acetylated α-pyrimidines) and purine nucleobases, one main observation has been made: acetylated sugars and nucleobases do not appear to react under the conditions employed. We hypothesise that this is because free hydroxyl groups at both the anomeric position and C-2 are required to properly co-ordinate Mg^{2+} and facilitate Lewis acid catalysis. We also investigated the reaction of other ribose derivatives that lack a 2-OH that is available for co-ordination by Mg^{2+}, including 2-deoxyribose **176** and arabinose *ara*-**30**, but no nucleoside products were observed in either case. This finding is corroborated by incubation of acetylated inosine derivatives under

glycosylation conditions: although deacetylation occurred, the glycosidic bond appeared to be completely stable. The reaction of further ribose derivatives (such as 2-methoxy ribose or per-methoxy ribose) under the same conditions could be investigated to further corroborate this hypothesis, or substitution of 2-OH for 2-SH (2-mercapto ribose) and Mn^{2+} substituted for Mg^{2+}. This hypothesis, however, does contradict Orgel's description of preliminary results, which indicate that 2-deoxyribose **176** and adenine **44** or hypoxanthine **153** do indeed form a glycosidic bond.

We also observed that ribose-5-phosphate *ribo*-**30**-5P does not appear to undergo glycosylation (in agreement with Orgel's findings[172]), but an explanation for this is unclear. It may be that phosphate provides a sufficiently active Lewis base to quench the activity of Mg^{2+} under the reaction conditions used, or has sufficient buffering capacity to prevent specific acid catalysis. Benner and colleagues, however, have recently succeeded in forming a glycosidic bond between phosphorylated ribose derivatives and nucleobases.[175] Ribose-1,2-cyclic phosphate was observed to react with adenine **44** and hypoxanthine **153** in the presence of divalent metal ions (Mg^{2+} or Ca^{2+}) when heated in dry state. The ß-nucleotide is formed in 8-15% yield: the cyclic phosphate blocks the lower face of the sugar, directing nucleobase attack to the top face; regioselectivity was also observed, reacting only at *N*-9 of adenine **44**, although it is not clear from where this selectivity is derived. The non-natural nucleotide 2′-phosphate formed can be selectively converted to the 5′-isomer when heated in the presence of inorganic phosphate, urea and borate (if borate is omitted from the reaction mixture, a complex mixture of products forms): the borate is suspected to co-ordinate the 2′,3′-cis diol and direct phosphate migration to the 5′-OH.

Ribose-1,2-cyclic phosphate itself can be formed from ribose under prebiotically plausible conditions, although ribose 2,3-cyclic phosphate is also produced. The two cyclic phosphate isomers are formed in a 3:2 ratio (with the 1,2-cyclic isomer as the major product) when ribose is reacted with amidotriphosphate (4 eq.) in the presence of $MgCl_2$ (29% yield of two isomers), or when reacted with DAP (1.2 eq., 71% yield).[153]

This glycosylation therefore shows a parsimonious recycling of reagents: borate, which has been reported to promote the formation of ribose by stabilising key intermediates in its formation under formose conditions,[147] is also required to efficiently rearrange 2′- to 5′-phosphorylated nucleotides; Mg^{2+} is required to form ribose-1,2-cyclic phosphate

from ribose and amidotriphosphate, and is also required for the reported glycosylation to proceed; DAP can also be used to form ribose-1,2-cyclic phosphate and has been implicated in other areas of prebiotic chemistry.[152–154] It therefore appears that direct glycosylation of ribose derivatives is not a prebiotic dead-end, and the reagents required unify this synthesis with other areas of prebiotic chemistry. Nevertheless, it is not clear if and how the parsimonious recycling of unwanted α-nucleotides, which initiated this investigation, can proceed.

5.1 Stereospecific reduction of cyclo-purines

5.1.1 Introduction

The high-yielding synthesis of pyrimidine ribonucleotides proceeding under prebiotically plausible conditions developed by Sutherland (*vide infra*) represents an important step towards elucidating the plausibility of RNA abiogenesis.[138] A complete prebiotic RNA synthesis, however, also requires the synthesis of purine ribonucleotides. At present, no analogous route of assembly has been revealed for the prebiotic synthesis of purines.

5.1.1 Routes to canonical purines

As discussed previously, the seemingly simple retrosynthetic analysis of purine ribonucleotides to phosphate, ribose, and a nucleobase is of little use in attempting a synthesis of purine ribonucleotides under prebiotically plausible conditions. The two canonical purine ribonucleotides, ß-adenosine ß-**55** and ß-guanosine ß-**56**, are produced in just 4% and 9% yield, respectively, from the dry-state heating of ribose *ribo*-**30** and the analogous free nucleobase in the presence of magnesium chloride and sodium trimetaphosphate (***Figure 5.1***).[172] The lack of regioselectivity observed in the purine nucleobase (which contains a number of reactive nitrogen atoms) and stereoselectivity in the sugar (leading to the formation of pyranosyl- and α-anomeric ribonucleotide products) means the natural ß-furanosyl product is formed only as part of a complex mixture of isomeric byproducts. Furthermore, this low yielding ribosylation is just of one step in the synthesis: ribose and the free purine nucleobases need to be formed first.

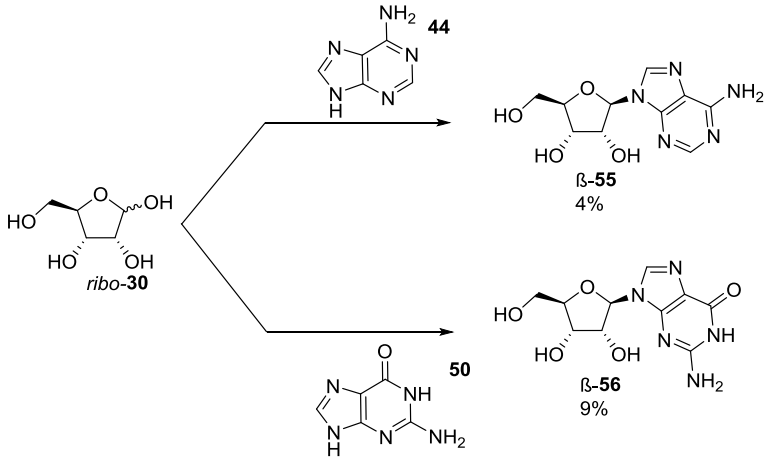


Figure 5.1*: Dry state heating of ribose ribo-**30** with adenine **44** or guanine **50** results in the formation of ß-adenosine ß-**55** and ß-guanosine ß-**56** respectively. Highest yields are achieved in the presence of magnesium chloride and trimetaphosphate.*

By manipulating competing glycosylation sites in the nucleobase moiety, Carell and co-workers have significantly increased the yield of purine ribonucleotides, observing up to 60% total yield of all adenosine isomers.[364] Instead of using free purine nucleobases, as Orgel and colleagues did,[172,348] Carell instead condensed ribose *ribo*-**30** and formamidopyrimidines (for example, **183**, ***Figure 5.2***).[364] These purine synthons can be formed from the reaction of formic acid **2** and aminopyrimidines (such as **184**), which in turn are formed from ammonium cyanide. The aminopyrimidines display proton-assisted reactivity guidance, with a proton effectively acting as a protecting group for most exocyclic amino groups, leaving the *N*-5 amino group the most nucleophilic. This then selectively reacts when the aminopyrimidine is heated in formic acid to give the *N*-5 formylated product.

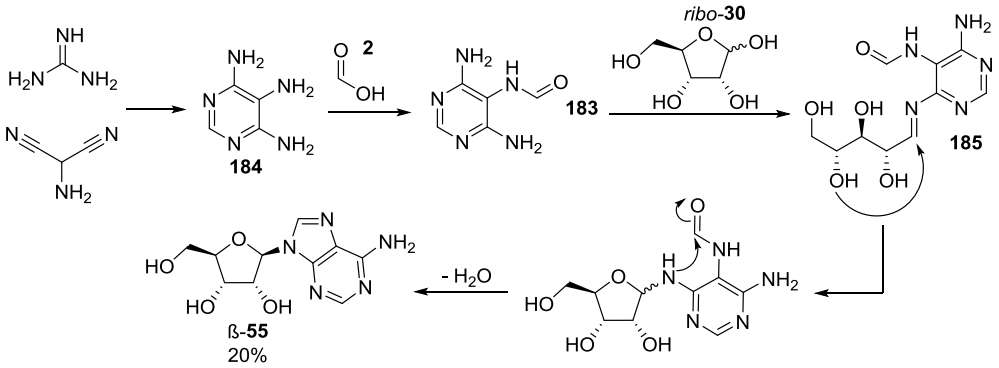


Figure 5.2*: A higher yield of ß-adenosine (ß-**55**, 20%) can be achieved by heating ribose ribo-**30** and formamidopyrimidine **183** in the dry state. Formamidopyrimidine can, in turn, be formed from the reaction of formic acid **2** and aminopyrimidines **184**.*

Condensation with ribose forms an imine **185**, which enables closure of the sugar ring, and a subsequent cyclisation provides the purine skeleton. This purine cyclisation step can, in fact, occur without the previous sugar-coupling taking place, meaning that the free nucleobase is formed as a by-product. Although Carrell's reaction represents an improvement in the regioselectivity of the sugar-nucleobase condensation, providing only *N*-9 ribosylated products, lack of selectivity in the sugar moiety during ribosylation results in the synthesis of a number of analogous byproducts (as well as the natural ß-furanosyl products, non-natural pyranosyl isomers and α-anomers are formed). Furthermore, this reaction is still reliant on the presence of ribose: the reaction of formamidopyrimidine with glycolaldehyde and glyceraldehyde (prebiotically plausible mixtures of sugars) in the presence of $Ca(OH)_2$ resulted in just 0.9% of ß-furanosyl-adenosine (the various purine pentosides were the main products, in around 21% yield, but the majority of the starting material is not accounted for).[364] The yield of ß-

furanosyl-adenosine increased to 1.3% when the reaction was carried out in the presence of borax.[364]

Pyrimidine and purine ribonucleotides would need to have appeared contemporaneously for functional RNA to be produced. While not strictly necessary, it would be highly advantageous for this complementary synthesis of purine ribonucleotides to take place under the same prebiotic conditions as those discovered for the selective formation of pyrimidines. One of the advantages of Sutherland's pyrimidine synthesis is its divergence to give uracil and cytosine in the final step of the synthetic route; a further divergence between the syntheses of pyrimidines and purines would also be advantageous.

One example strategy of divergent prebiotic nucleotide synthesis, involving early tethering of the nucleobase precursor to the sugar precursor has been explored by Powner and co-workers (***Figure 5.3***).[365] This couples the aminooxazoline chemistry used in the first steps of the previously established pyrimidine synthesis with the HCN chemistry of prebiotic nucleobase synthesis, circumventing the requirement for a glycosylation reaction and ensuring the purine is attached at the correct *N*-9 position. When AICA **48** (the hydrolysis product of AICN, itself a intermediate in the polymerisation of HCN to give adenine) is mixed with 2-aminooxazole **26** and glyceraldehyde **33** at pH 5, a three-component coupling reaction occurs: **48** and **33** first react to generate imine **186**, which is trapped by 2-aminooxazole to give a masked C-5 sugar **187** with an imidazole tethered at C3′. Ring closure then results in the attachment of the imidazole moiety at C1′ (this bond corresponds to the glycosidic bond in the nucleoside) of the three-component product **188**, which is generated in 84% total yield (by ^{1}H NMR integration) as a mixture of diastereomers. This product contains a C5-sugar backbone with the correct oxidation level at each carbon, tethered at the correct C1′-position to the correct nitrogen atom (analogous to *N*-9 in the canonical product) of a partially constructed nucleobase. The analogous product is also formed when AICN **45** is used instead of AICA **48**, but competing 6-*exo*-*trig* imidazole cyclisation and 5-exo-trig hydroxyl cyclisation result in a 1:1 mixture of products. The *lyxo* product dominates (60% yield by ^{1}H NMR integration), which would establish the correct stereochemistry at C1′, C3′ and C4′ after C3′-stereoinversion and further elaboration to provide the ß-ribo-adenosine 2′,3′ cyclic phosphate nucleotide.

Furthermore, this reaction shows a pH dependent distribution of pyrimidine-precursor and potential purine-precursor products.[365] At pH 7, even in the presence of excess AICA and AICN, these purine precursors do not react. Instead, two-component chemistry occurs, providing the pentose aminooxazolines **62** as a mixture of diastereomers, with *ribo*-**62** (44%) and *ara*-**62** (30%) dominating: *ara*-**62** can then be elaborated to give the canonical cytosine and uracil nucleotides.[138] In this way, products of increasing complexity can be provided in a selective manner under potential prebiotic conditions.

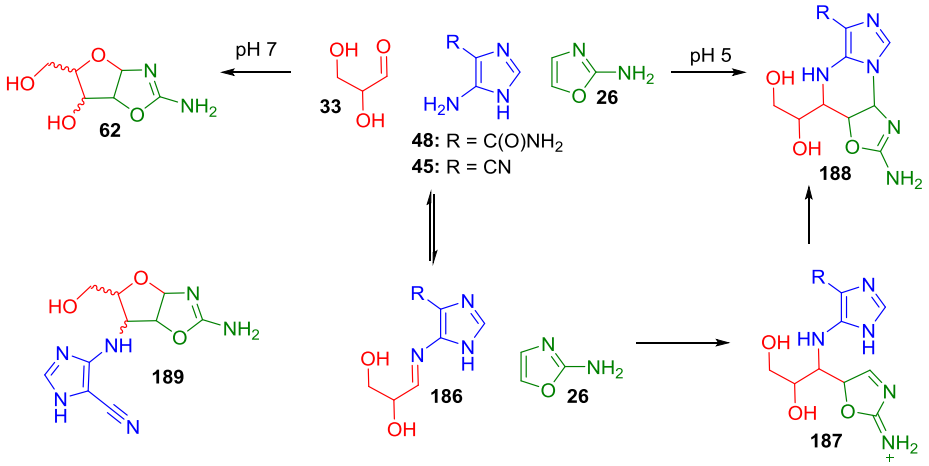


Figure 5.3: *AICA* ***48*** *or AICN* ***45,*** *glyceraldehyde* ***33*** *and 2-aminooxazoline* ***26*** *react in a pH dependant manner, affording different products at pH 7 and pH 5. At pH 7, two-component chemistry occurs between* ***33*** *and* ***26****, giving the pentose aminooxazolines* ***62****. At pH 5, three-component chemistry takes place:* ***33*** *and* ***48*** *or* ***45*** *react to form an imine* ***186****, which undergoes nucleophilic attack by* ***26****. Ring closure yields the three-component product* ***188*** *as a mixture of diastereoisomers in 84% total yield. An additional product,* ***189****, is seen when AICN* ***45*** *is used.*

5.1.2 A route to 8-oxo purines

It is clear that a synthetic approach to canonical purine synthesis using a tethered purine precursor can overcome the glycosidation problem, but the optimal tethering site to direct stereo- and regioselective-purination is unclear. Recent work in the Powner group has established a divergent synthesis of anhydro-pyrimidines and anhydro-purines from a shared prebiotic thione precursor which allows the circumvention of unstable free sugars and direct nucleobase glycosidation.[342] Both classes of anhydro-nucleoside undergo phosphorylation and stereoinversion to yield nucleotides with the correct ß-*ribo* sugar stereochemistry: the natural pyrimidines, cytosine **31** and uracil **51**, and oxidised analogues of the natural pyrimidines, 8-oxo-adenosine **190** and 8-oxo-inosine **191** (***Figure 5.4***).

It has previously been demonstrated that glycolaldehyde **32** and cyanamide **4** react to form 2-aminooxazole **26**, which already contains the future glycosidic bond; subsequent reaction with glyceraldehyde **33** forms the aminooxaolines as a mixture of isomers, with the desired *ribo*- and *arabino*- isomers dominating.[138] The reaction of glycolaldehyde and thiocyanic acid **192** (a product of the reaction of hydrogen cyanide and sulfur) generates an analogue of 2-aminooxaole, 2-thiooxazole **193**. This thione can be crystallised directly from water and transported by sublimation, suggesting that prebiotic mechanisms for its purification, transportation and accumulation were accessible on the early Earth. Reaction with glyceraldehyde **33** at neutral pH and at 60°C generates a mixture of the pentose oxazolidinone thiones (**194a**, R=H) in good yield (74%) after 24 h with the sulfur selectively positioned at C2. Only the *lyxo* diastereomer is present as a mixture of the furanosyl and pyranosyl isomers; complete natural furanosyl selectivity is seen in the other three diastereomers, with high *ribo/arabino* diastereoselectivity observed (70% *ribo/arabino* 1:1, 30% *lyxo/xylo*). The *ribo*- and *arabino*-thiones are each only a single stereoinversion from the natural ß-ribofuranose sugar found in RNA. Cyanoacetylene **5** can be used to selectively activate the thione, providing the S-cyanovinylthione (**194b**, R = CHCHCN) in quantitative yield after just 1 h; subsequent addition of methanethiol **195** provides the more stable S-alkyl thione (**194c**, R = Me) in 50% yield over two steps.

At this point in the synthetic route, a second (and final) point of divergence to give pyrimidines and 8-oxo-purines occurs: thione displacement with ammonia from *arabino*- or *ribo*-cyanovinyl thiones provides the respective aminooxazolines *ara*- and *ribo*-**62**, which can be elaborated to give the canonical pyrimidines. Hydrogen cyanide oligomer **196**, meanwhile, is able to substitute either the *S*-cyanovinyl thione or the *S*-alkyl thione to give the aminooxazolines **197** in good yield (81% when R=Me from the alkyl thione) and subsequently undergo 5-*exo-dig* cyclisation to build the imidazole moiety of the purine nucleobases on a sugar scaffold. Unlike the previously established prebiotically plausible route towards pyrimidines, in which divergence to give the natural pyrimidines occurs in the final step, here, divergence towards the different purine analogues occurs at a relatively early point in the synthesis. The purine precursor is tethered by a 2′,8-anhydro bond which provides a method for generating the natural ß-ribo sugar by phosphorylation and inversion at C2′. Purine elaboration is achieved by heating (100°C) the aminoimidazoles in a mixture of the hydrogen cyanide derivatives formamide and formamidine: 2′,8-anhydroadenosine *ara*-**145** was observed

to formed in 65% yield after 5 h, but the (non-canonical) 2′,8-inosine derivative was produced in much lower yield (just 11% after 48 h). Urea-mediated 2′,8-anhydroadenosine *ara*-**145** phosphorylation results in C3′-OH phosphorylation and inversion at C2′, generating the natural ß-ribo sugar with a 2′,3′-cyclic phosphate moiety, which, by virtue of its inherent ring strain, activates the ribonucleotide to oligomerisation.

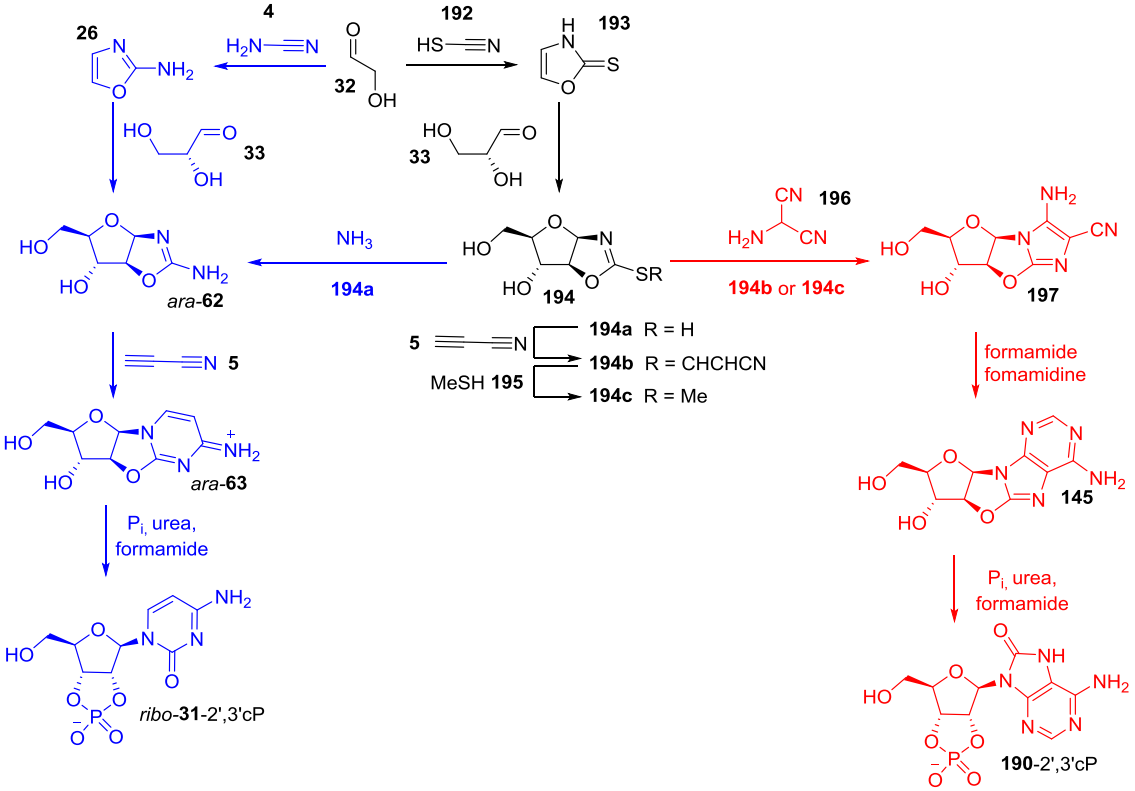


***Figure 5.4**: Divergent synthesis of pyrimidines (blue) and 8-oxo-purines (red). Both classes of products contain a 2′,3′-cyclic phosphate group, which is inherently activated towards oligomerisation. Formation of the cyclic phosphate moiety results in inversion at C-2′ and breaking of a 2′-nucleobase anhydro bond, generating the nucleotide in the natural ß-ribo configuration.*

Although this synthetic route is noteworthy in generating 8-oxo-purines and canonical pyrimidines from a common precursor, it does not generate canonical purines. The route from **190** to an activated form of a canonical nucleotide (**55**-2′,3′cP) is apparently simple, requiring only a single reduction.

Previous work in the Powner group has also established that 8,2′-anhydro-8-adenosine *ara*-**145,** under weakly alkaline conditions, undergoes intramolecular rearrangement to give 8,5′-anhydro-8-oxyadenosine *ara*-**146**,[366] and under strongly

alkaline conditions, intramolecular rearrangement to the epoxide **147** (***Figure 5.5***).[345] If the activation inherent to 8,2′-anhydro-8- adenosine *ara*-**145** could be exploited twice and the two steps from *ara*-**145** to *ara*-**146** and *ribo*-**190** combined, then 8,5′-anhydro-8-oxy-2′,3′-cyclic adenosine monophosphate *ribo*-**146**-2′,3′cP would be produced. A reduction of the 8,5′-anhydro bond would then provide a canonical nucleotide with an activated cyclic phosphate, **55**-2′,3′cP.

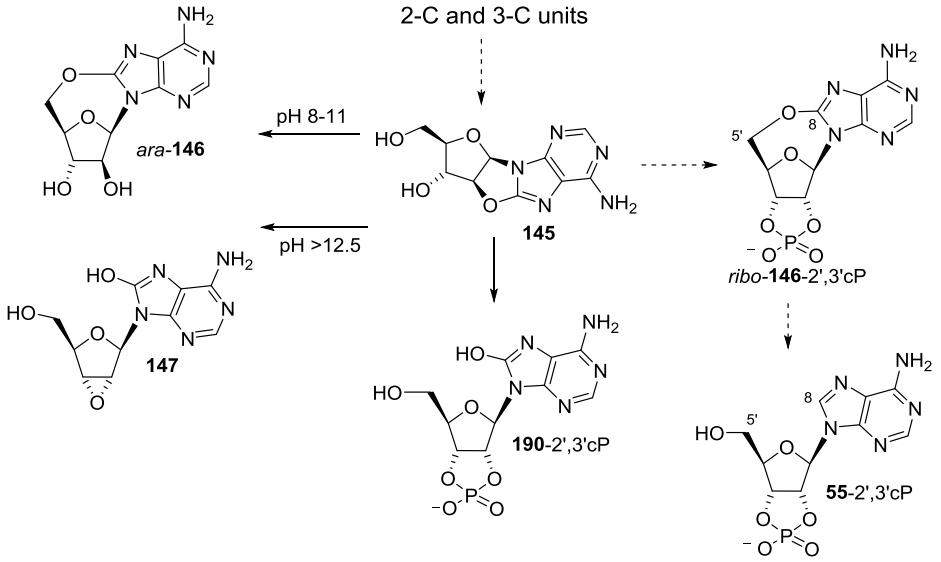


Figure 5.5*: Previous work in the Powner group on the synthesis of canonical adenosine nucleotides, and proposed route via reduction to* ***55****-2',3'cP. Ara-****145*** *undergoes intramolecular rearrangement under alkaline conditions to give the 8,5'-anhydro derivative ara-****146*** *(pH 8-11) or the epoxide* ***147*** *(pH >12.5), and stereoinversion at C2' upon phosphorylation to generate* ***190****.*

5.2 Project aims

It has been established in the literature that 8,5′-anhydro-8-oxyadenosines (eg **198**) are acid-labile, undergoing C(5′)-O bond cleavage to give the corresponding 8-hydroxyadenosine (**199,** ***Figure 5.6a***).[367,368] Conversely, Maki *et al*. have reported a highly selective reduction of the C(8)-O bond of *N*-6-acylated nucleoside **200** mediated by sodium cyanoborohydride in acetic acid (***Figure 5.6b***), but non-acylated 8,5′-anhydro-8-oxy-2′,3′-*O*,*O*-isopropylidene adenosine **198** was observed to be stable under these reaction conditions.[369] Maki and colleagues propose the *N*-6-acyl group is required for the reductive cleavage of the O-C(8) bond, suggesting that the acyl group coordinates a proton, facilitating *N*-7 protonation, facilitating addition of hydride at C8.

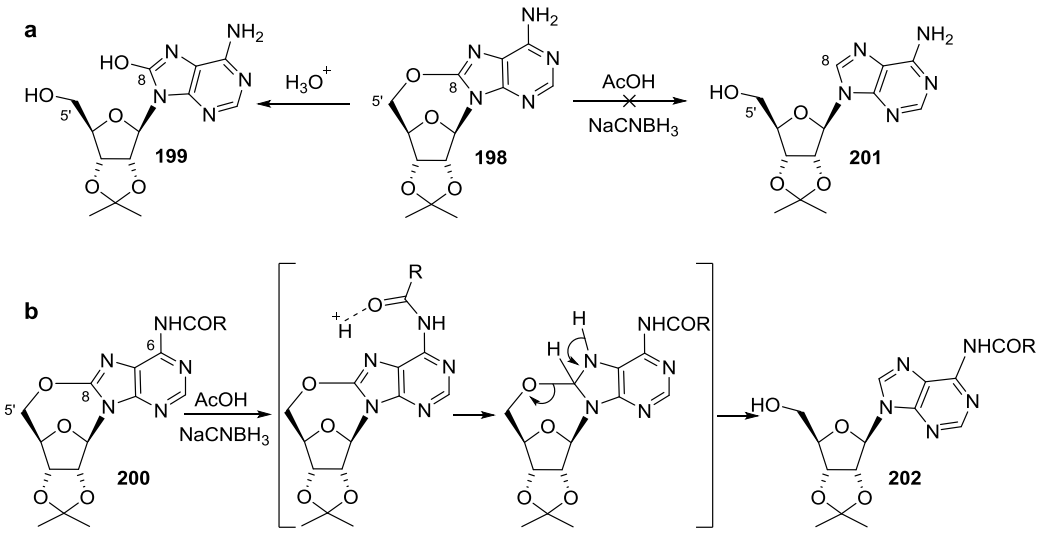


Figure 5.6: *Ring-opening reactions of 8,5′-anhydro-8-oxyadenosines.*

a: *2′,3′-O,O-isopropylidene-8,5′-anhydro-8-oxyadenosine* ***198*** *undergoes acid-mediated hydrolysis to give the 8-hydroxy nucleoside* ***199****, but not reductive cleavage of the 8,5′-anhydro-8-oxy-ring.*

b: *The presence of an N-6 acyl group in* ***200*** *allows facile reductive cleavage of the C(8)-O bond.*

The conformational changes imposed on the sugar ring in a ribonucleotide that has its 2′- and 3′-hydroxyl groups linked by a five-membered ring (such as an isopropylidine protecting group or a cyclic phosphate) has been widely reported in literature. This bicyclic system confers an unusual rigidity on the ribose ring[370], meaning it adopts ‘east’ or ‘west’ structures that are otherwise rarely seen.[138] The bicyclic system leads to a partial flattening of the chair structure and it becomes more envelope-like: four ring atoms tend towards planarity, with only O-4′ easily puckered.[370,371] In the ‘west’ conformation, the 1′ and 4′ substituents of the sugar ring are brought closer together which provides a high effective molarity of the C-1′ substituent and the C-4′ substituent (***Figure 5.7***).[372,373] The presence of a 2′,3′-isopropylidine protecting group therefore promotes an increased rate of bond formation between the 5′-OH and pyrimidine nucleobases.[374,375] Similarly, ß-cytidine-2′,3′-cyclic phosphate is believed to be unusually stable (compared to the α-anomer and cytidines without a 2′,3′-cyclic phosphate) when subjected to UV irradiation because 5′-OH is held in close proximity to the nucleobase and so can form a stable photo-adduct containing an anhydro-bond.[138]

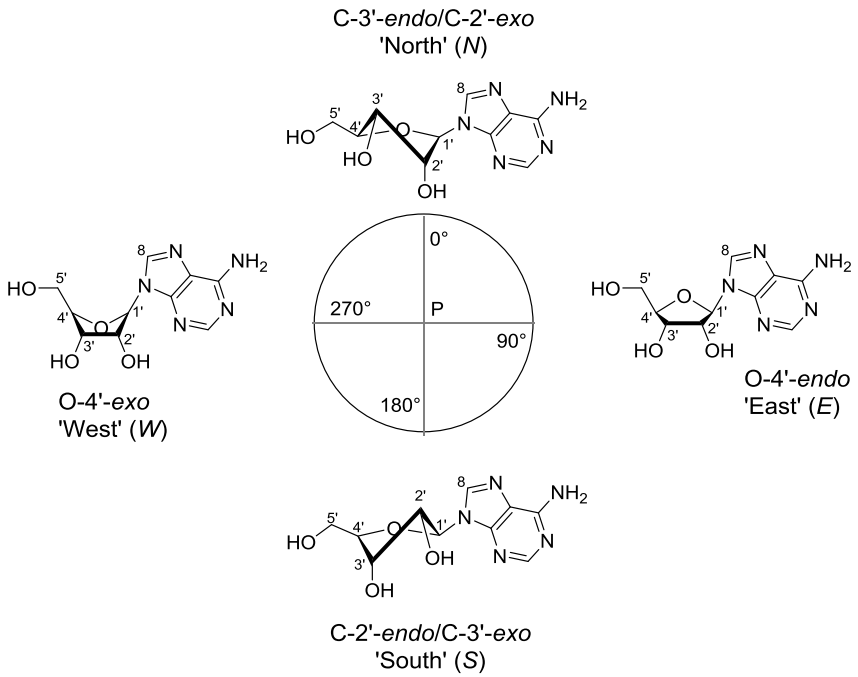


Figure 5.7: *The pseudorotation wheel describes a continuum of ring puckers seen in furanose rings of nucleotides (here, adenosine is shown): P (the phase angle of pseudorotation) is defined in terms of the five torsion angles of the furanose ring.*[376,377] *When one ring atom is out of the plane of the other four, the ring adopts an envelope structure (such as O-4'-endo 'East' or O-4'-exo 'West') – but these structures are rarely observed as it is energetically more stable for the ring substituents to be held as far apart as possible.*[377] *The furanose ring usually adopts a structure in which two ring atoms are out of plane with the other three: ribonucleotides usually adopt a C-3'-endo ('North') conformation, and deoxyribonucleotides take on a C-2'-endo ('South') conformation.*[377] *The geographical descriptions of the sugar conformations relate to the position of the P values for each conformation on the pseudorotation wheel.*[377]

As the isopropylidine protecting group used by Maki and colleagues induces the same structural conformation as a 2′,3′-cyclic phosphate – both bringing the 5′-OH and C-8 into close proximity – it is possible that this type of reduction (using either sodium cyanoborohydride, or a more prebiotically plausible reductant) may also occur in **146**-2′,3′cP (or, if a proton co-ordinating group is indeed necessary, an *N*-6 acylated derivative). If nucleobase protonation is essential to reduction, we hypothesised that the zwitterionic nature of the first protonated conjugate acid of **146**-2′,3′cP would make it more receptive to reduction under conditions where proton transfer could occur (such as in water).

The principal aim of this project was to discover a prebiotically plausible reduction of **146**-2′,3′cP to give the canonical nucleotide **55**-2′,3′cP. Initial investigations of this reduction would first take place using the non-prebiotically plausible **198**, an analogue of **146**-2′,3′cP, with well-documented chemistry, and using known non-prebiotically plausible reductants. If these attempts were successful, then the prebiotic formation of **146**-2′,3′cP and the efficiency of prebiotically plausible reductants would be explored.

5.3 Results and Discussion

5.3.1 Synthesis and reduction of 2′,3′-*O,O*-isopropylidene-6-*N*-benzoyl-8,5′-anhydro-8-oxyadenosine *203*

The work of Maki *et al*. was first replicated to prove that this reduction would work as published,[369] which necessitated the synthesis of a suitable substrate (***Figure 5.8***). Maki and co-workers attempted the reduction of both **202b** and **202a** (the *N*-6-benzoyl and *N*-6-acetyl derivatives, respectively), with the former giving the highest yields: it was therefore decided that this proof-of-concept study should use the benzoyl derivative. 2′,3′-*O*,*O*-Isopropylidene-6-*N*-benzoyl adenosine **202b** – synthesised by the benzoylation of commercially available 2′,3′-*O*,*O*-Isopropylidene-adenosine **201** - was isolated in 62% yield. Reaction of **202b** with *N*-iodosuccinamide in acetic acid[378] afforded the desired 5′,8-anhydro product **203b** in 58% isolated yield. An initial reduction with $NaCNBH_3$ in acetic acid was successful, with 70% conversion to the desired product **202b** estimated from ^{1}H NMR integral analysis.

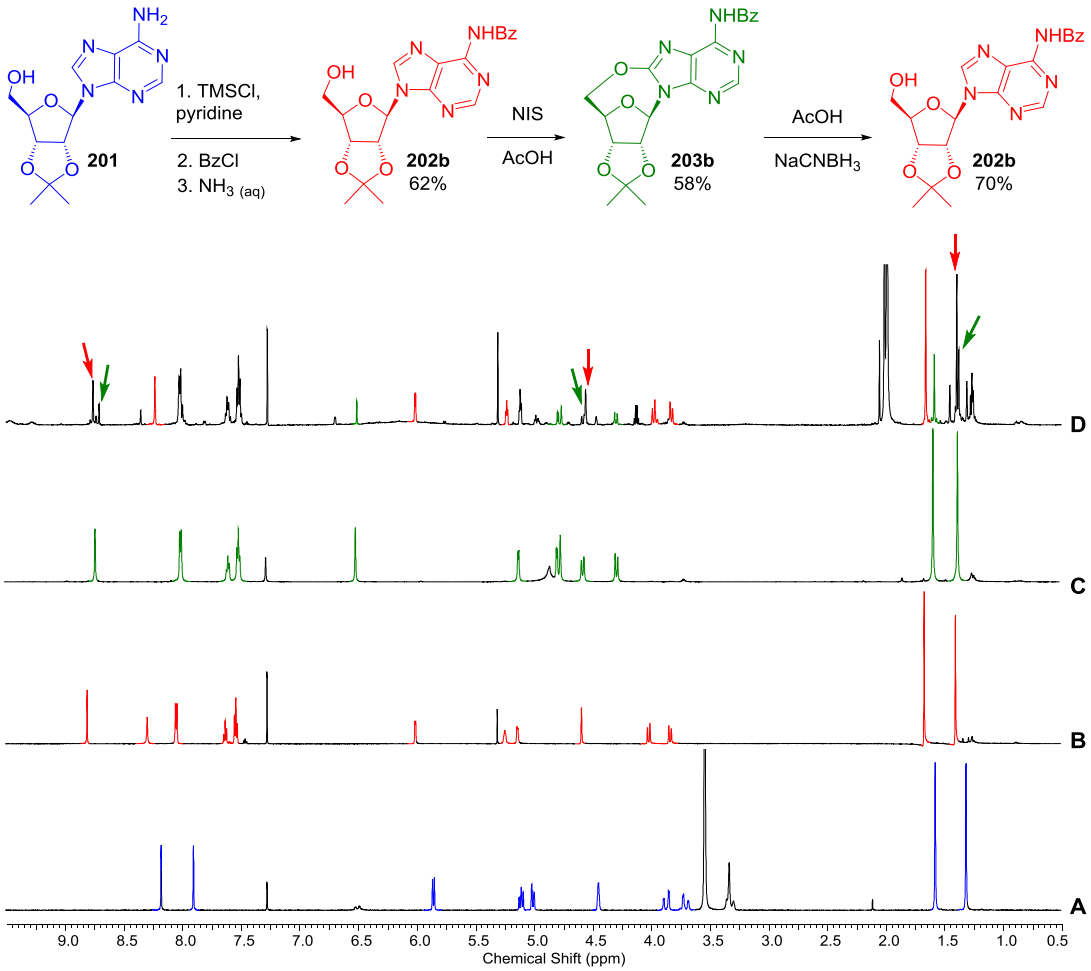


Figure 5.8: *Synthesis, and subsequent reduction, of* ***203b****.* ***Top****: Scheme showing synthetic route to, and reduction of* ***203,*** *under the conditions described by Maki et al.* ***Bottom****: ^{1}H NMR spectra (600 MHz, 0.5-9.5 ppm, $CDCl_3$ unless specified) comparing* ***A)*** *commercial* ***201****, highlighted in blue ($CDCl_3$ and CD_3OD)* ***B) 202b****, highlighted in red;* ***C) 203****, highlighted in green; and* ***D)*** *the crude product of the reduction of* ***203b*** *with $NaBH_3CN$ (7 eq.) after 25 h at r.t. in AcOH and subsequent water quench. The starting material (****203b****, highlighted in green) has undergone incomplete conversion to the reduction product (****202b****, highlighted in red).*

Reductions were also attempted using $NaBH_3CN$ in TFA or formic acid, and formic acid as solvent and reductant, but the desired product could not be detected by NMR analysis in the crude products of these reactions.

Although Maki *et al*. reported that this reduction did not occur without the presence of an *N*-6 acyl group, they do not provide any details of variation of the reaction conditions attempted, if any, to drive the reaction. Therefore, 2′,3′-*O*,*O*-isopropylidene-8,5′-anhydro-8-oxyadenosine **198** was synthesised and its reduction attempted.

5.3.2 Synthesis of 2′,3′-*O*,*O*-isopropylidene-8,5′-anhydro-8-oxyadenosine *198*

A number of syntheses of **198** exist in literature. Guo and co-workers have reported a simple one-step procedure, using iodine and $NH_3 \cdot H_2O$ at 60 °C, achieving a yield of 34% (***Figure 5.9***).[379] We found that **201** was not fully soluble in the reaction solvent; adaptions made to improve the reaction mixture resulted in disappointingly low isolated yields (maximum 6%) of **198**. (The presence of an *N*-6 benzoyl group coordinates a proton, facilitating the protonation of *N*-7, rationalising the increased yields seen in in this reaction with such nucleoside derivatives).

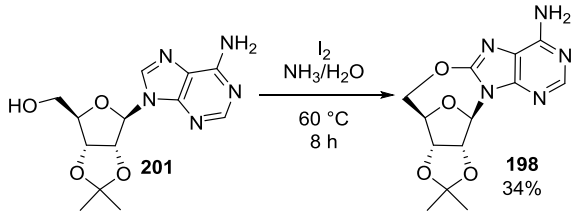


Figure 5.9: *Guo et al.'s synthesis of* ***198*** *using iodine and ammonia.*

Maki and co-workers' simple one-step procedure, using *N*-iodosuccinamide in acetic acid, achieved only an isolated literature yield of 24% with this particular nucleoside analogue[378] (lower than Guo's reported yield), but our two attempts using this procedure were unsuccessful.

A two-step synthesis of **198** *via* 2′,3′-*O*,*O*-isopropylidene-8-bromoadenosine **204** was developed in the late 1960s (***Figure 5.10***).[367,380,381] This procedure involves bromination of **201** in a mixture of dioxane and disodium hydrogen phosphate solution, and purifying crude **204** by recrystallization for use in the next step.[380] **204** is then stirred with NaH in dry dioxane overnight, unreacted NaH quenched, the mixture filtered, the filtrate evaporated and recrystallized, and the crystals obtained combined with those isolated from the first filtration, providing **198** in 64% yield. [367,381]

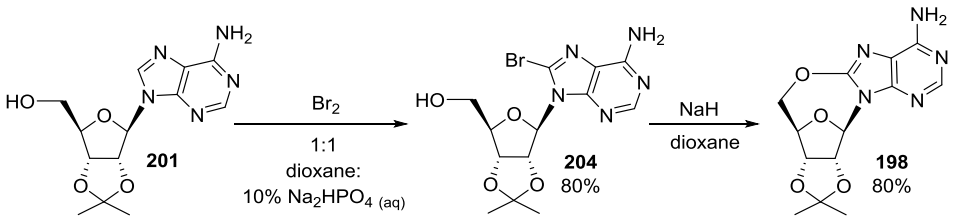


Figure 5.10: *Scheme showing Ikehara's two-step synthesis of* ***198****, via the C-8 brominated product,* ***204****. Bromination of C-8 in basic solution followed by cyclisation under anhydrous conditions results in a literature yield of 64% over two steps. Potter has recently achieved a slightly higher yield (83%) of* ***204*** *using this protocol.*[382]

We hypothesised that the cyclisation step could take place in aqueous solution at high pH (with the 5′-OH deprotonated by hydroxide, instead of hydride, allowing it to attack the electrophilic C-8), offering a two-step process with yields uncompromised by increased simplicity. It was found that **198** could be attained this way in 60% overall yield (***Figure 5.11***): comparable to Ikehara's yield over two steps, but a significantly simpler operational protocol.

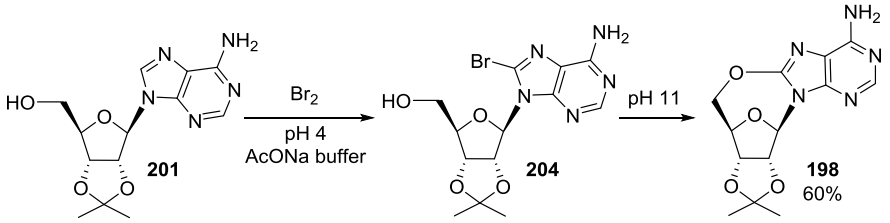


Figure 5.11: *Scheme showing facile one-pot synthesis of* ***198****. After bromination in aqueous solution at pH 4, cyclisation could be triggered by raising the pH of the reaction mixture, achieving a yield of 60% after purification by crystallisation over two steps.*

The successful establishment of a synthetic route to **198** meant that its reduction, and reduction following further derivitisations, could be investigated.

5.3.3 NMR identification of 8,5′-anhydro-8-oxyadenosines

The absence of ^{1}H NMR coupling and 2D NMR COSY correlation to the proposed anomeric proton (H-1′, δ 6.33) indicates that this molecule has a strained conformation, featuring dihedral angles that disfavour $J_{1',2'}$ coupling. The H-5′ protons also undergo a characteristic downfield shift, moving from around δ 3.8 to 4.2 ppm. These changes, which allow easy monitoring of the progression of the reaction, can be easily seen in ***Figure 5.12***. Similar changes were also observed in the *N*-6 benzoylated derivatives.

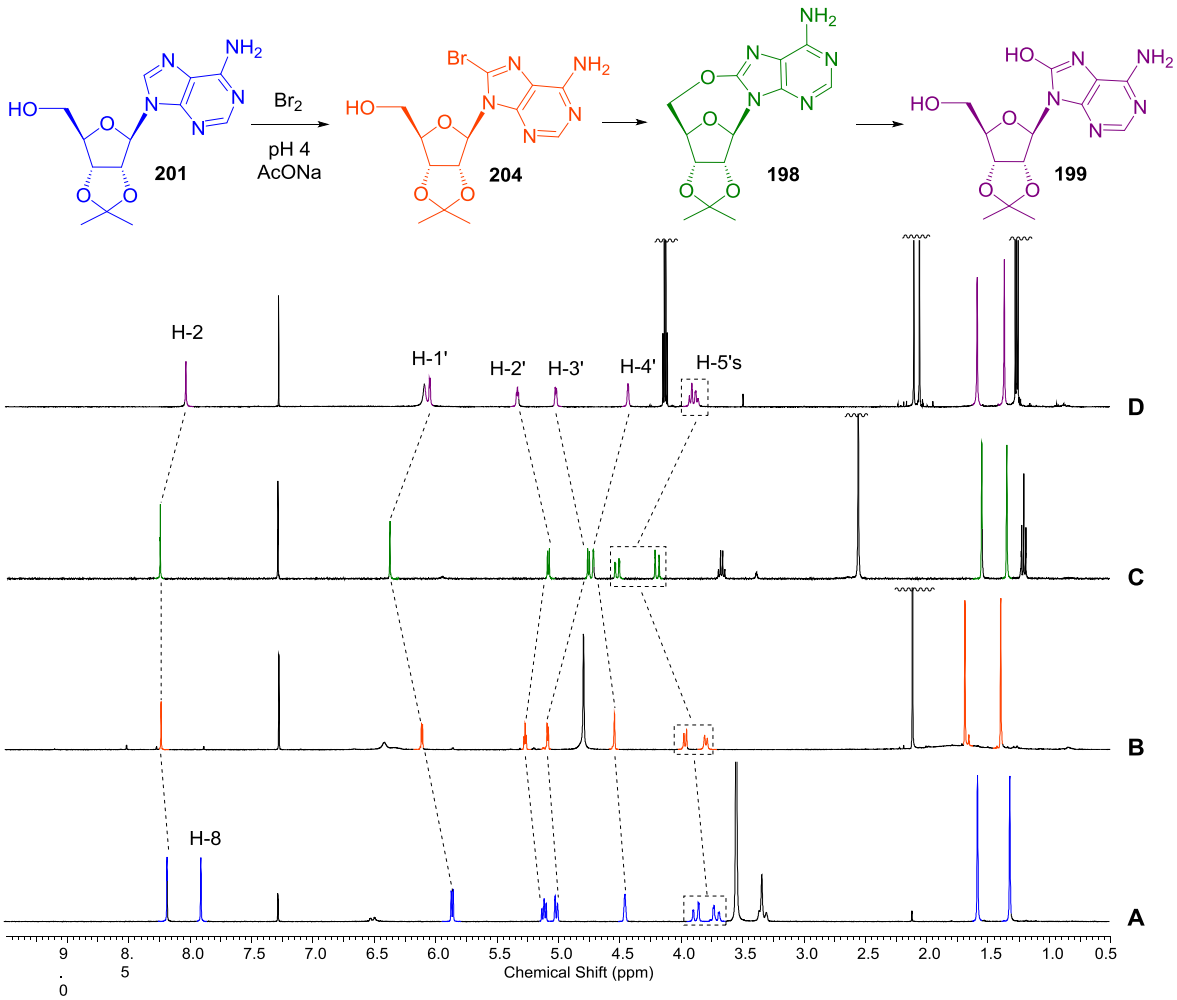


Figure 5.12: *^{1}H NMR identification of adenosine derivatives. Comparison of sections of ^{1}H NMR spectra (600 MHz, 0.5-9.5 ppm,* ***A:*** *$CDCl_3$ and MeOH;* ***B,C,D****: $CDCl_3$). An H-8 peak can only be seen for starting material* ***201*** *(blue) – all other species are substituted at this position. A slight downfield shift of H-1' can be seen in the* ***204*** *(orange). A significant downfield shift of H-1' and H-5's signify the formation of the 5',8-anhydro ring. The H-1' resonance for* ***198*** *is a singlet, but a doublet in other species.*

Although many papers reporting the synthesis and reactions of 8-cyclopurines have not published full NMR spectroscopy data,[367,368,383–386] the anomeric proton appearing as a singlet peak seems to be a known feature of 2′,3′-isopropylidine-8,5′-cyclo-purine nucleosides (both 8,5′-*O*- **198** and 8,5′-*S*-cycloadenosine **205** in the isopropylidene protected forms display the H-1′ signal as a singlet [387,381,388,389]) in contrast to 8,2′ and 8,3′-cyclopurines. [390,391]

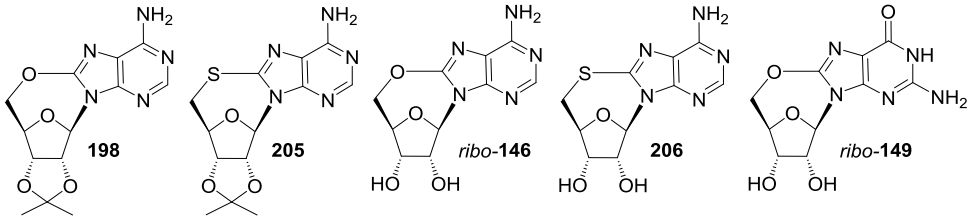


Figure 5.13: *Structures of a selection of 8,5′-cyclopurines.*

It is unclear from literature, however, whether this is also characteristic of 8,5′-cyclopurines without extra ring strain conferred by a cyclic 2′,3′-protecting group. In a review of his work, Ikehara states that both 8,5′-*O*- and 8,5′-*S*-cycloadenosine, *ribo*-**146** and *ribo*-**206** (in the non-isopropylidene protected forms) show no coupling at H-1′ (no NMR data is provided in this review for 8,5′-cyclo guanosines or the isopropylidene protected forms of 8,5′-cyclo adenosines);[391] in a later paper, he reports the H-1′ signal as a singlet in 8,5′-*O*-cycloadenosine *ribo*-**146** and 2′,3′-isopropylidine-8,5′-*O*-cycloadenosine **198**.[381] A subsequent paper reports ^{1}H NMR data for unprotected cyclopurines: *ribo*-**206** has a H-1′ singlet; whilst *ribo*-**146** has a H-1′ doublet, although, curiously, its coupling constant is 0 Hz.[388] More recently, full NMR data for *ribo*-**146** in DMSO-d_6 has been published by both Yu[392] and Capon,[393] who both report the same chemical shifts for ^{1}H NMR. Although Yu has not assigned any NMR data, Capon attributes a singlet peak at δ 5.98 to the anomeric proton. A paper by Kadlubar *et al*. reports the 1′-H signal of 8,5′-*O*-cycloguanosine *ribo*-**149** as a doublet (but does not provide a coupling constant);[386] perplexingly, they follow the synthesis of Kohda *et al*. who present the change of the 1′-H signal into a singlet as proof that they have successfully formed the 8,5′-anhydro bond.[385] Data from other members of the Powner group suggests that 8,5′-cyclopurines with free 2′- and 3′-OHs show coupling at H-1′, but a diol deprotection of **198** to provide *ribo*-**146** in this project gave a species with a H-1′ singlet (*vide infra* section 5.3.5).

5.3.4 Attempted reductive cleavage of 2′,3′-*O*,*O*-isopropylidene-8,5′-anhydro-8-oxyadenosine *198*

The general method of Maki *et al*. was adapted in an attempt to find conditions under which a cyclo-adenosine without an *N*-6 acyl group (**198**) could undergo reduction.[369] Under the standard conditions ($NaBH_3CN$ in acetic acid) no reduction took place. After an extended reaction time (69 h) with a total of 31 eq of $NaBH_3CN$, 35% conversion to an unidentified product had taken place; many of its other peaks overlap with those of **198** and other species present making its identification difficult.

When TFA was substituted for AcOH, **198** underwent complete conversion to two deprotected derivatives. The ^{1}H NMR data of the initially-formed species suggests that it contains a 5′,8-anhydro bond, but also that it is not 8,5′-anhydro-8-oxy-adenosine *ribo*-**146.** This unidentified species converted to a second nucleoside without a 5′,8-anhydro bond after further reaction time.

Ikehara and co-workers have reported that 5-*O*-8-adenyl-ribose **207** is formed as a byproduct of the acid hydrolysis of **198**, along with their desired product *ribo*-**146**, and **208** and 8-oxyadenine **209** (***Figure 5.14a***).[367,381] Comparison with the partial literature ^{1}H NMR data available indicates that the unidentified product first formed is not α- or ß-**207**, both of which have a very upfield-shifted anomeric proton which couples to H-2′.[394] It is possible the unidentified species first formed is 8,5′-anhydro-8-oxo-7-(ß-D-ribofuranosyl)adenine **210**, in which case the second unidentified species could be 8-oxo-7-(ß-D-ribofuranosyl)adenine, **212**. (Literature ^{1}H NMR data indicates that 7-(ß-D-ribofuranosyl)adenine **211** has not been formed;[395] furthermore only one only one singlet peak in the aromatic region appears to integrate to the anomeric proton, indicating that the C-8 position is substituted.)

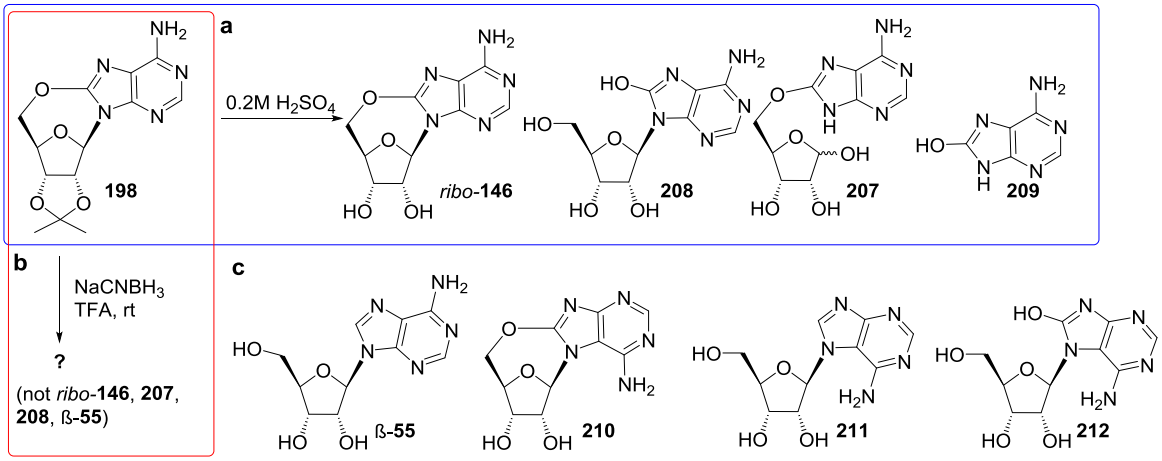


Figure 5.14: reaction of 198 with acid

a*: Ikehara's observations with H_2SO_4 (blue box)*

b*: Reaction with TFA/$NaCNBH_3$ in this project (red box)*

c*: Structures of adenosine ß-**55**, 8,5'-anhydro-8-oxo-7-(ß-D-ribofuranosyl)adenine **210**, and 7-(ß-D-ribofuranosyl)adenine **211** and 8-oxo-7-(ß-D-ribofuranosyl)adenine **212.***

The lack of success in the attempted reduction of 2′,3′-*O*,*O*-isopropylidene-8,5′-anhydro-8-oxyadenosine **198** under a variety of conditions was disappointing. The products formed could not be identified, but it was clear that (at least when TFA was used as a solvent) they had undergone diol deprotection.

5.3.5 Attempted diol deprotection of 2′,3′-*O*,*O*-isopropylidene-8,5′-anhydro-8-oxyadenosine *198*

It has been reported that cyclisation *via* nucleophilic attack of 5′-OH on 8-C can only take place if the nucleoside also contains a 2′,3′-cyclic group:[381] this holds the 5′- and 1′-substituents of the sugar closer together. The same structural conformation has been used to justify the reaction between a 5′-substitutent and a nucleobase, forming an anhydro-bond, in a variety of pyrimidine nucleotides (discussed in further detail in the aims of this chapter).[138,373–375] It is plausible that if **198** is deprotected, the 8,5′-anhydro ring would be more strained and therefore could be broken more easily. The 5′,8-anhydro ring is known to be acid labile, but there is literature precedent for diol deprotection using acid with retention of the anhydro bond.[381,392] The deprotection of **198** was accordingly attempted in using sulfuric acid,[381] acetic acid, and TFA.[392]

Analysis of the ^{1}H NMR spectrum of the deprotection by sulfuric acid was complicated by the presence of contaminants – probably from Dowex resin used to work up the reaction. Nevertheless, analysis by mass spectrometry proved that the desired product was present (***Table 5.1***, entry 1). Substitution of acetic acid for sulfuric acid afforded the desired product, 5′,8-anhydro adenosine *ribo*-**146**, in 47% yield after 32 h, alongside two further diol-deprotected species with H-1′ doublet peaks (***Table 5.1***, entry 2). It is possible that these are 8-oxo derivatives of adenosine. Our attempts to replicate a literature deprotection of **198** using TFA in a refluxing mixture of water and methanol were unsuccessful (***Table 5.1***, entry 3): no change took place after the stated reaction time, but only 8-oxo-adenosine **208** was obtained when the reaction was refluxed overnight (***Table 5.1***, entry 4).[392] In various concentrations of dilute TFA, **198** underwent very little conversion (***Table 5.1***, entries 5-8), but both the desired product and the hydrolysis product were observed each reaction when analysed by ^{1}H NMR.

The latter attempt (entry 8: using TFA in water 9:1) appeared promising so was repeated, and the reaction progress followed by removing the solvent *in vacuo* at various time points and subjecting the crude residue to NMR analysis. On this attempt, no hydrolysis product was observed: instead, **198** was partially (and cleanly) deprotected, with 65% conversion seen after 5.5 h.

Entry	Reaction conditions	Relative integration of species displaying a H-1′ peak observed by ^{1}H NMR		
		198	*ribo*-**146**	**208**
1	0.5 M H_2SO_4, 60 °C, 5h	-	-[a]	-
2	0.75 M acetic acid, 60 °C 22 h, then 90 °C 10 h.	33%	47%	-[b]
3	2:1:1 MeOH:TFA:H_2O, reflux, 5 h	100%	-	-
4	2:1:1 MeOH:TFA:H_2O, reflux, overnight	-	-	100%
5	TFA:H_2O 1:3, rt, 5 h	80%	4%	16%
6	TFA:H_2O 1:1, rt, 5 h	72%	6%	22%
7	TFA:H_2O 3:1, rt, 5 h	80%	6%	14%
8	TFA:H_2O 9:1, rt, 5 h	19%	70%	11%

Table 5.1: *Attempted deprotections of* ***198*** *using various acids.*

[a] *detected by mass spectrometry*

[b] *two unidentified species detected (in 7% and 13% yield), one of which could be* ***208****.*

These initial studies implied that 2′,3′-*O*,*O*-isopropylidene-8,5′-anhydro-8-oxyadenosine **198** could be deprotected under acidic conditions with retention of the delicate 5′,8-anhydro bond. Further studies would be required to optimise this deprotection and develop a purification of *ribo*-**146**. The reductive cleavage of *ribo*-**146** could then be attempted, or a cyclic phosphate group attached at the 2′,3′-diol to provide 8,5′-anhydro-8-oxy-2′,3′-cyclic-adenosine monophosphate **146**-2′,3′cP, which could be more receptive to reductive cleavage. **146**-2′,3′cP could, of course, be synthesised by more direct methods.

5.3.6 Synthesis of 8,5′-anhydro-8-oxy-2′,3′-cyclic-adenosine monophosphate *146-2′,3′cP*

Although we have shown that **198**, the isopropylidine-protected analogue of nucleotide **146**-2′,3′cP, could be easily synthesised in aqueous conditions, the inevitability of 2′,3′-cyclic monophosphate hydrolysis in basic conditions whilst attempting to form the 8,5′-anhydro bond in the cyclic nucleotide meant a different approach would be required. We envisaged the best way to form this bond would be by adapting Ikehara's procedure discussed earlier.[381] Cyclising the nucleotide with NaH in anhydrous solvent would, however, require the nucleotide to be soluble (or at least partially soluble) in organic solvent. Therefore, a bulky organic counter ion on the charged cyclic phosphate would be required.

8-Bromo nucleotide **213**-2′,3′cP was formed by reaction with bromine-water in pH 4 acetate buffer (***Figure 5.15***). The next step in our envisaged route (ion exchange of Na^+ to a bulky, organic-soluble counter ion) first required rigorous purification to remove the acetate buffer. **214**-2′,3′cP was successfully isolated, but in just 4% yield over two steps from **55**-2′,3′cP.

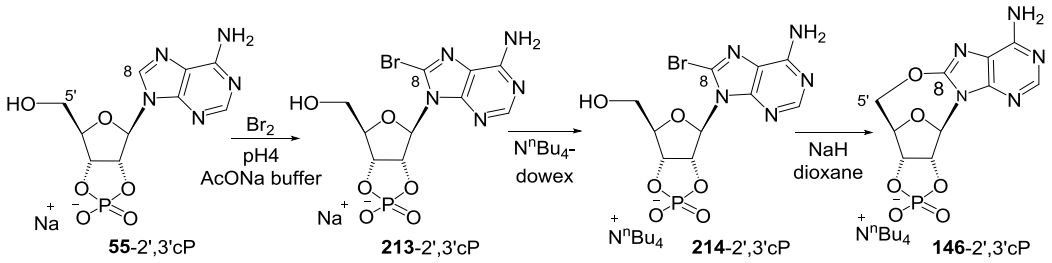


Figure 5.15: *Proposed synthesis of 8,5′-anhydro-8-oxy-2′,3′-cyclic-adenosine monophosphate* ***146****-2′,3′cP.*

As **55**-2′,3′cP was not conveniently available**,** further attempts to synthesise **214**-2′,3′cP, with improved yield, were instead made with the 2′/3′-mono phosphate mixed isomers **55**-2′P and **55**-3′P, which necessitated the addition of an extra phosphate-cyclisation step to the synthesis. We found the best yields could be achieved by bromination of the mixed isomers first, followed by ion exchange then phosphate cyclisation (***Figure 5.16***). Following this protocol, **214**-2′,3′cP was obtained in 35% yield over 3 steps, with purification by flash column chromatography carried out just once, at the end of the synthetic route.

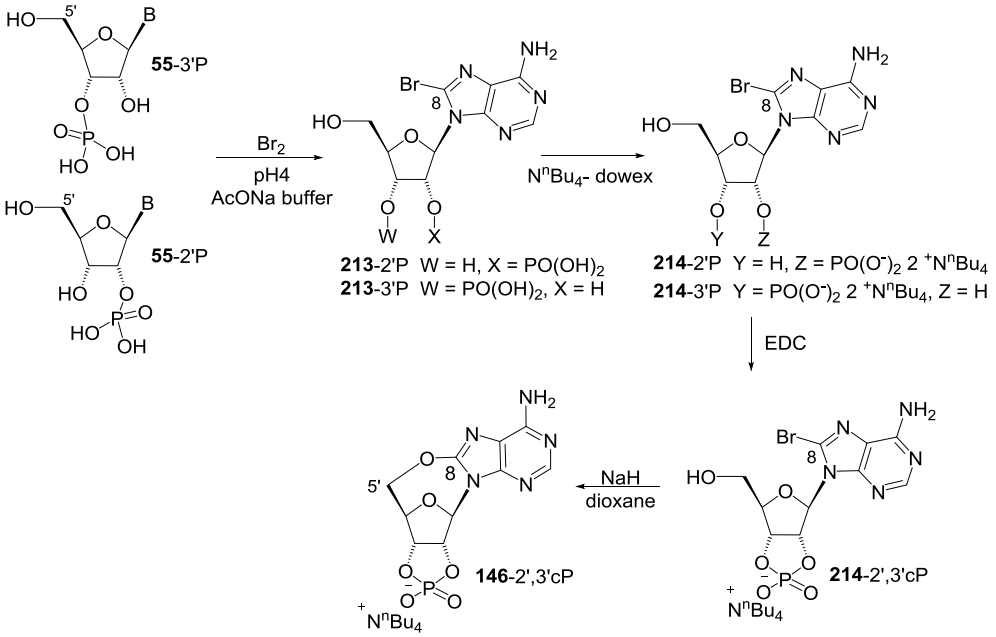


Figure 5.16: *Revised route to* ***146****-2’,3’cP from regioisomers* ***55****-2'P and* ***55****-3'P (B=A).*

With **214**-2′,3′cP in hand, we next investigated its 5′,8-cyclisation, but to our dismay **214**-2′,3′cP was very unsoluble in dioxane, the solvent we originally conceived the 5′-8-cyclisation would be carried out in. **214**-2′,3′cP was slightly more soluble in ^{t}BuOH, so this was used with NaH and ^{t}BuOK as bases instead. Unfortunately, **146**-2′,3′cP was not observed to form: these reaction conditions instead resulted in the opening of the 2′,3′-cyclic phosphate.

5.3.7 Attempted synthesis of 8,5′-anhydro-8-oxy-guanosines

Until this point, exploratory work on the oxidation of purines to form a 5′,8-anhydro-8-oxy product followed by reduction had been carried out using adenosines: adenosine is the easiest purine to handle, as it is more soluble in both organic and aqueous solvents.[396] Guanosine **56**, however, is the most easily oxidisable base as it has the lowest redox and ionisation potentials.[384] A number of procedures for the synthesis of a 5′,8-anhydro bond in guanosine (and guanosine derivatives) exist in literature.[384–386] The reaction mechanisms of these syntheses share a common principle: cyclisation occurs following oxidation of guanosine, with cyclisation occurring by attack of the 5′-hydroxyl at the positively charged or polarised C8 atom.

Kohda report that 8,5′-anhydro guanosine *ribo*-**149** is formed from the reaction of guanosine **56** and DNPA **215** (***Figure 5.17***), albeit as a minor component in a complex mixture of products;[385] this method has been used more recently for the preparation of *N*-7-amino guanosine **216**.[386] This reaction was attempted to test the potential of optimising this route to 8,5′-anhydro guanosine *ribo*-**149**. A complex mixture of products was indeed formed, and moreover, the precise nature of many of these products were not apparent (the anomeric proton of the desired product, for example, has been reported as a doublet by Kadlubar[386] and as a singlet by Kohda[385]). It was therefore decided to focus on other routes to 8,5′-anhydro guanosine.

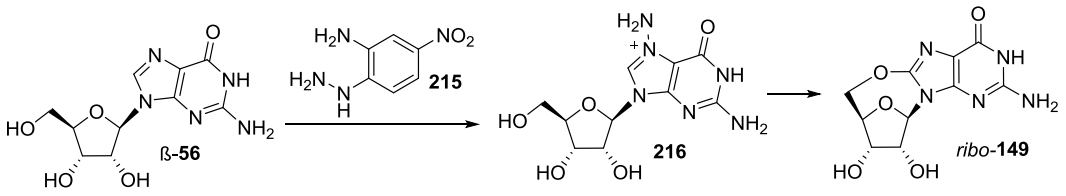


Figure 5.17: *Attempted synthesis of 8,5′-anhydro guanosine ribo-**149** with DNPA **215**.*

Choi *et al.* have investigated a number of oxidations of deoxyguanosine monophosphates (the DNA guanine derivative) using a Pt^{IV} complex, tetraplatin, **217**.

The authors found that, in the presence of tetraplatin, deoxyguanosine 5′-monophosphate **218**-5′P forms 8-oxo-3′-deoxy guanosine-5′-monophosphate **219**,[397] but when deoxyguanosine 3′-monophosphate **218**-3′P is used as a substrate 8,5′-anhydro-8-oxy-3′-deoxy-guanosine monophosphate **220** is instead formed (***Figure 5.18***).[384]

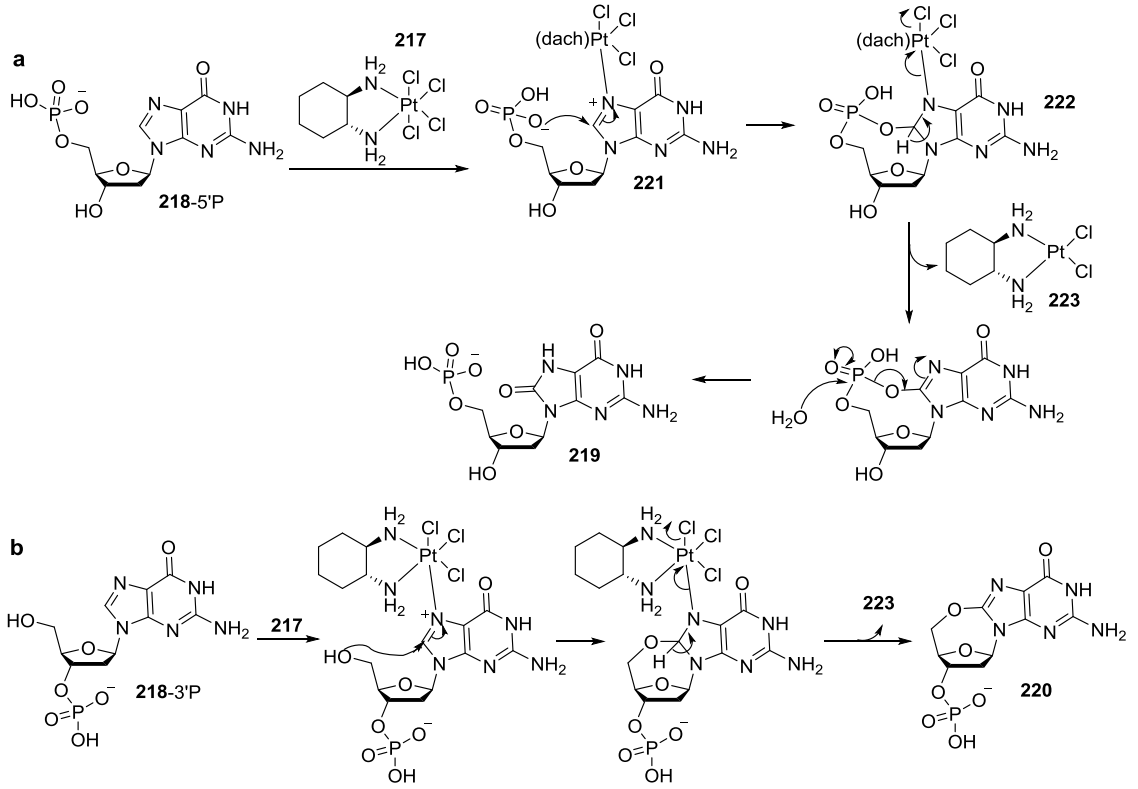


Figure 5.18*: Reactions of **217** with deoxyguanosine monophosphates. **217** binds to N-7 of the guanosine derivative, with subsequent attack of a C-5′ nucleophile on C-8.*

a*: 5′-dGMP **218**-5′P forms the 8-oxo derivative **219***

b*: 3′-dGMP **218**-3′P forms the 8,5′-anhydro derivative **220***

*(**217**, tetraplatin = trans-d,l-1,2-diaminocyclohexanetrachloroplatinum, [$Pt^{IV}Cl_4$(dach)]; dach = diaminocyclohexane)*

The proposed redox mechanisms for both these products are initially the same: a chloride ion is lost from tetraplatin **217**, which then binds to *N*-7 of the guanosine derivative.[384,397] Tetraplatin **217** has a high reduction potential[384] and efficiently withdraws electron density from C-8 when bound to *N*-7, allowing a C-5′ nucleophile to attack the polarised C-8. If this intramolecular nucleophile is a hydroxyl group, [$Pt^{II}Cl_2$(dach)] and a nucleotide containing a 5′,8-anhydro bond, **220**, are released.[384] In contrast, attack by a 5′-phosphate group results in the formation of a cyclic phosphodiester **221**; and a nucleotide containing a cyclic phosphodiester (**222**) and

[$Pt^{II}Cl_2$(dach)] **223** are released.[397] The cyclic phosphodiester is less stable than the cyclic ether, and therefore hydrolyses to afford the 8-oxo derivative, **219**.[397] The 8-oxo-derivative can also be obtained when the sugar moiety of guanosine is replaced by an ethyl group, and the reaction carried out in the presence of aqueous inorganic phosphate or hydroxide: the intermolecular nucleophile attacks the polarised C-8 of the platinated guanine derivative and de-protonates this carbon, followed by two-electron transfer releasing the 8-phosphate (which then hydrolyses) and 8-hydroxyl derivatives.[398] Further evidence for the intramolecular addition of phosphate was acquired when the reaction of deoxyguanosine 5′-monophosphate was carried out in $H_2{}^{18}O$, which showed that only 1% of the final product was obtained from direct attack of water on the platinated intermediate, and demonstrating that the intramolecular phosphate oxyanion is a better nucleophile.[397]

Choi's reaction with deoxyguanosine 3′-monophosphate **218**-3′P was replicated.[384] The appearance of a new species in the ^{1}H NMR spectrum (H-1′: δ 6.43, d, J = 6.1 Hz, ***Figure 5.19***) corresponds to the observations of Choi *et al*., indicating that 3′-deoxyguanosine monophosphate **218**-3′P had partially converted to 8,5′-anhydro-8-oxo-3′-deoxyguanosine monophosphate **220** (24% by relative integration of anomeric protons after 6 h). The cyclisation of a number of analogues was also attempted. The reactions of both guanosine phosphate mixed isomers (**56**-2′P and **56**-3′P) and guanosine **56** with tetraplatin **217** produced a complex mixture of products when analysed by ^{1}H NMR, and the respective desired products could not be identified in these mixtures.

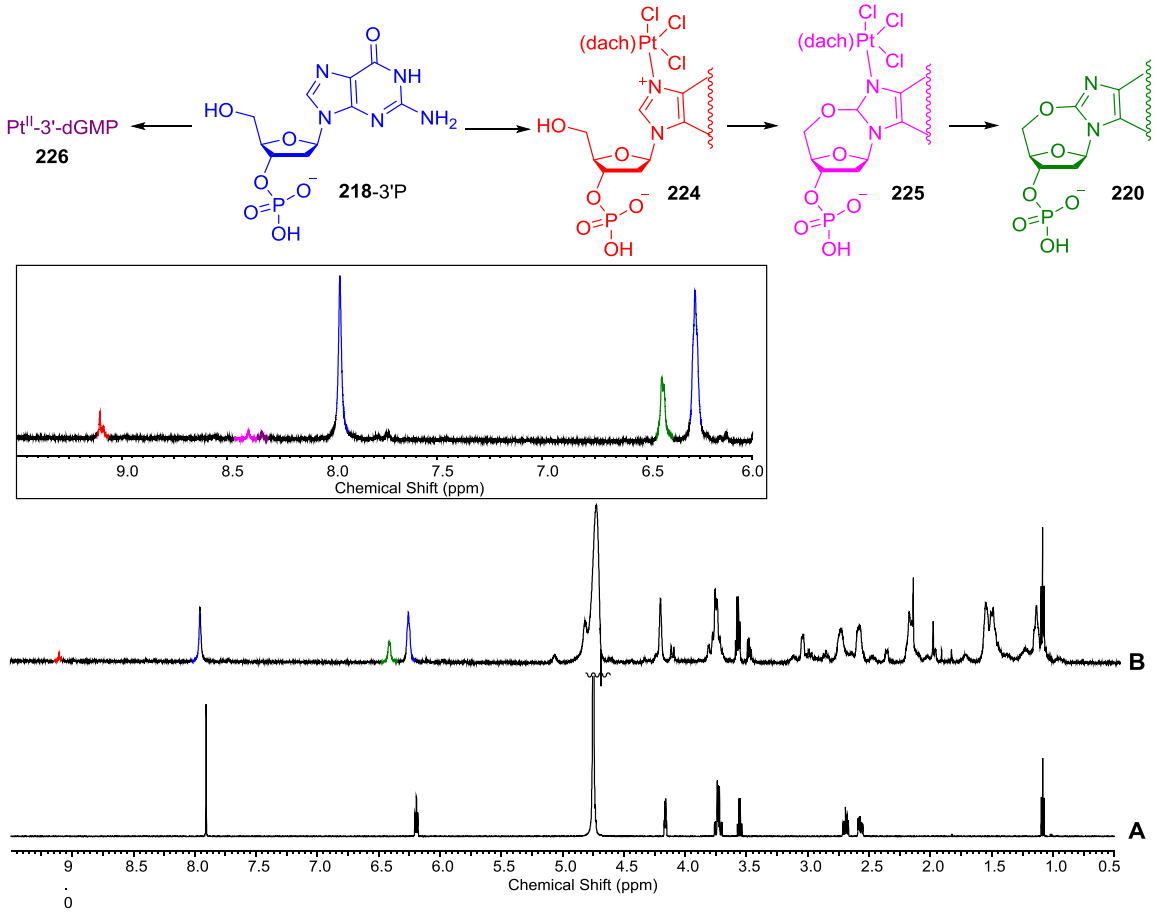


Figure 5.19**: Reaction of deoxyguanosine 3′-monophosphate **218**-3′P with **217.

***Top**: **218**-3′P reacts with **217** to give the Pt^{IV}-N7 intermediate (**224**, red). Attack of the 5′-OH species at C-8 results in the formation of the Pt^{II}(5′,8-anhydro) intermediate (**225**, pink). Loss of [Pt^{II}Cl$_2$(dach)] **223** releases 8,5′-anhydro-8-oxo-3′-deoxyguanosine monophosphate (**220,** green); [Pt^{II}Cl$_2$(dach)] can then react with free **218**-3′P to provide Pt(II)-3′dGMP (**226**, purple).*

***Bottom**: ^{1}H NMR spectra (400 MHz, D_2O, 0.5-9.5 ppm) showing the reaction of deoxyguanosine 3′-monophosphate **218**-3′P with **217**. **A**) Commercial deoxyguanosine 3′-monophosphate; and **B**) after reaction with **217** for 6 h. Inset: section of the ^{1}H NMR spectrum showing reaction of deoxyguanosine 3′-monophosphate with **217** after 6 h.*

2′,3′-Cyclic guanosine monophosphate **56**-2′,3′cP (which is not commercially available) was synthesised from **56**-2′P and **56**-3′P using EDC **22** and its cyclisation then attempted (***Figure 5.20***). As before, ^{1}H NMR analysis of the reaction mixture proved that it increased in complexity, although a ^{31}P NMR of the reaction mixture after 27 h shows that only 2′,3′-cyclic phosphate is present. Assuming the cyclic phosphate group confers the same structural rigidity on the ribose ring as an isopropylidene protecting group does on the ribose ring of 8,5′-cyclo adenosine, the desired product should be identifiable by a singlet in the H-1′ region of the ^{1}H NMR. A singlet peak does appear in

the ^{1}H NMR at a reasonable chemical shift: however, relative integral analysis of the H-1′ region reveals it represents only 27% of all species present. Furthermore, there is very little change in conversion after the first 2 h of reaction time; this corresponds to a dramatic fall in pH from 8.8 to 4.1 after 2 h, with very little further change in pH after this.

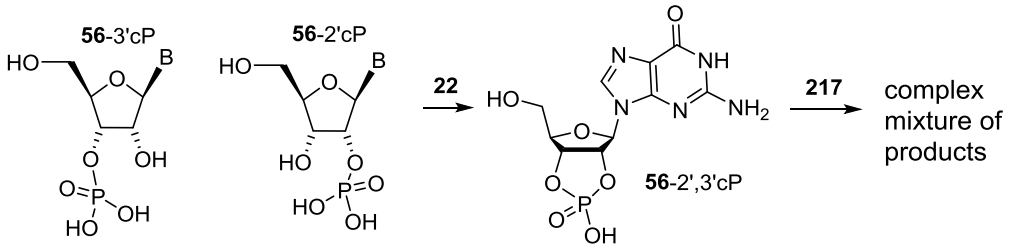


***Figure 5.20**: Synthesis of 2',3'-cyclic guanosine monophosphate **56**-2',3'cP and reaction with tetraplatin **217** (B = G)*

Therefore, the same reaction was attempted with the pH manually controlled so that it did not fall below 7.6. This resulted in very little conversion (approximately 10% of the species present in the H-1′ region), but no cyclic phosphate hydrolysis.

The variety of products generated with each substrate, and lack of corroboration in literature as to how the desired products should appear by ^{1}H NMR, made each of these reactions difficult to analyse. This was compounded by problems with the solubility of **217**: although Choi's recommendations to dissolve **217** over several hours and repeated heating/vortexing cycles before adding the nucleotide were followed,[384] precipitation still occurred throughout the reaction. It was therefore decided that this synthetic route should be discontinued for the present.

5.4 Conclusions and outlook

Limited progress has been made towards investigating the possible reduction of 5′,8-anhydro nucleotides over the course of this project, in part because of the difficulties encountered in synthesising these key intermediates. A facile one-pot two-step synthesis and purification of 2′,3′-*O*,*O*-isopropylidene-8,5′-anhydro-8-oxyadenosine **198** was attempted, but unfortunately attempts to find conditions for the reductive cleavage of **198** were unsuccessful. Initial investigations into the diol deprotection of **198** appear promising, but will require further optimisation and a method of purification developed. It is possible that 8,5′-*O*-cycloadenosine *ribo*-**146** is more susceptible to reductive cleavage than its isopropylidiene protected counterpart.

It may be necessary to uncover alternative means for synthesising 5′,8-anhydro nucleotides on large enough scale to explore their reduction. Nevertheless, some of the routes already attempted could be adapted. For example, 8-bromo-2′,3′-cyclic adenosine monophosphate **214**-2′,3′cP could be synthesised via the route already developed, and its conversion to 8,5′-anhydro-8-oxy-2′,3′-cyclic adenosine monophosphate **146**-2′,3′cP attempted in another solvent (such as pyridine). Furthermore, if a derivative of **214**-2′,3′cP was made with a water-soluble counter ion, its activity in aqueous solution could be attempted. Although this would almost certainly result in the cleavage of the cyclic phosphate group, it is possible that the 5′,8-anhydro bond could be formed under conditions similar to those used for the isopropylidene derivative.

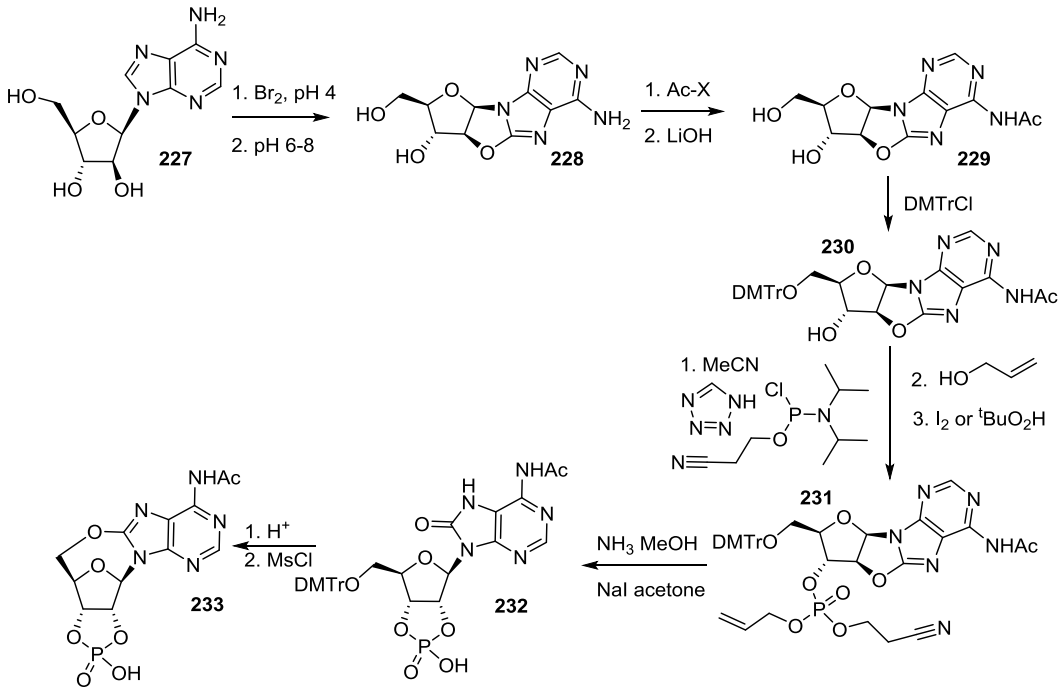


Figure 5.21: *Possible organic synthetic route towards* ***233****, an N--6 acylated derivative of ribo-****146****-2',3'cP.*

The reduction of an *N*-6-acylated derivative may prove more promising (prebiotically plausible methods for the acetylation of nucleotides exist in literature). There is no requirement, however, for the synthesis of this derivative to be restricted to prebiotically plausible conditions: via a synthetic organic route, such as that shown in ***Figure 5.21*** may be more efficient.

Once 8,5′-anhydro-8-oxy-purines have been synthesised in sufficient quantities, their hydride- and sulfide-mediated reductions (using dithiothreitol, for example) could be investigated. These reductions could also be investigated on 8,2′-anhydro-8-oxy-purines (such as *ara*-**145**, one of the final intermediates in the previously established prebiotically plausible synthesis of 8-oxo-adenosine).

Other reactions of 8,5′-anhydro-8-oxy-purines can also be explored. It has previously been reported that 2′,3′-O,O-isopropylidene-8,5′-anhydro-8-oxyadenosine **198** can react with a protected uridine 3′-phosphate **234** in anhydrous DMF (***Figure 5.22***).[368] The reaction proceeds by attack of the nucleotide phosphate anion on the 5′-C of the anhydro linkage, resulting in the formation of a dinucleoside phosphate with a natural 3′,5′-phosphate linkage (**235**). If such a reaction could be found under prebiotically plausible conditions, it would allow the formation of nucleotide oligomers from an intermediate of the synthesis of the nucleotides themselves.

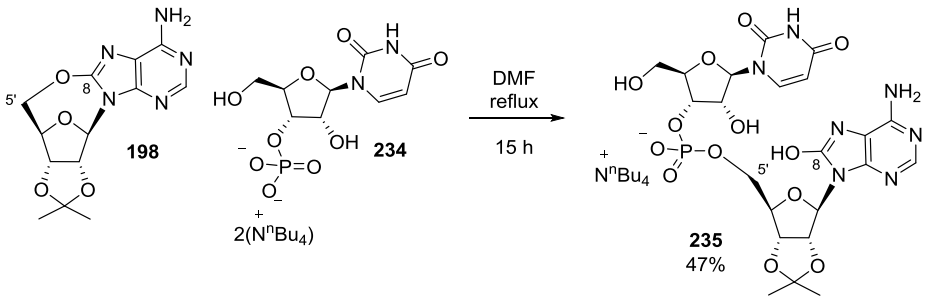


Figure 5.22: *Synthesis of a dinucleoside phosphate,* ***235****, from* ***198*** *and* ***234*** *under anhydrous conditions.*

Nevertheless, it must be kept in mind that the formation of 8,5′-anhydro-8-oxy-purines under prebiotically plausible conditions is, as yet, still hypothesised and has yet to be achieved. Furthermore, it may not even be necessary for the 5′,8-anhydro bond to be reduced at all: it is known that 8-oxo-purines, although not canonical, have been shown to be incorporated into RNA and tolerated during replication.[342]

6. Conclusion

Through our research into prebiotic chemistry, we have we have established a route to amino acid precursors, at neutral pH, that installs synthetic handles that could be used to allow further reaction and oligomerisation. The route to amino acid precursors that we have demonstrated here solves some problematic incompatibilities between the different reaction conditions required for the generation of different prebiotically-relevant molecules. We demonstrate links between this synthesis and the generation of other classes of metabolites through the use of a common reagent, and these links indicate our new potentially prebiotic building blocks are worthy of further investigation. This demonstration of successful transformations opens new avenues to be explored and narrows the focus of future prebiotic research. Equally important are the reactions we have demonstrated do not work in the prebiotic synthesis of nucleotides, which further constrain the boundaries of future research and highlight, by elimination of unsuccessful routes, areas upon which focus should instead lie. These studies of the feasible routes towards the biomolecules required for life refine our knowledge of the prebiotic chemistry that led to life. There are many plausible prebiotic pathways towards the molecules required for life, but many more pathways that will lead to dead ends;[399] discovery of the dead ends is nevertheless required to chart the chemistry of the prebiotic Earth.

We have established that DAP is an effective ammonia synthon in the neutral pH Strecker synthesis of amino nitriles (***Figure 6.1***). This phosphoro-Strecker reaction takes place at neutral pH – meaning it could potentially take place in the same environment as previously established successful nucleotide syntheses – unlike the conventional Strecker synthesis with ammonia, which requires more basic conditions that would block the synthesis of RNA by anhydronucleotide inversion[138] or thiocytidine photoanomerisation.[128] The integration of genetic information and proteins in life today indicates the importance of evaluating the coevolution of nucleic acids and peptides, and suggests a unification of nucleotide and amino acid syntheses is required. Demonstrating a route towards amino acid precursors that resolves this pH incompatibility between the conventional Strecker synthesis and plausible prebiotic nucleotide syntheses is a significant advance towards elucidating the unified origins of nucleic acids and peptides.

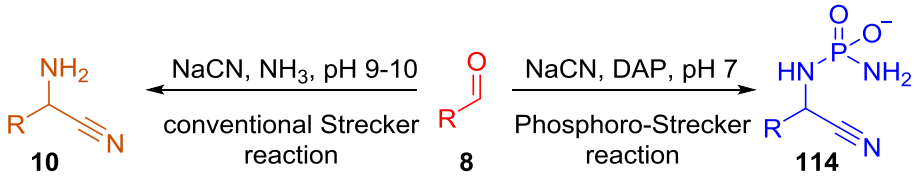


Figure 6.1: *The Strecker reaction at basic pH (left), generating an aminonitrile* ***10****, and the phosphoro-Strecker reaction at neutral pH (right), generating* ***114****.*

Moreover, we have observed that a pH switch operates when both DAP and ammonia are present: at pH 10, conventional Strecker reaction to give the aminonitrile predominantly occurs, whereas at pH 7 the phosphorylated derivative instead forms (***Figure 6.2***). Phosphate acts as a universal energy currency in life today, notably activating amino acids for protein synthesis before tRNA charging, so the selective formation of a phosphorylated amino acid derivative is particularly interesting, demonstrating how phosphate and peptide chemistry can be united.

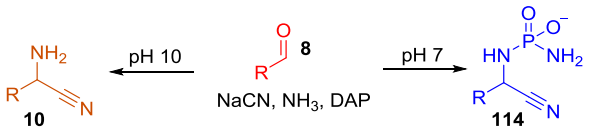


Figure 6.2: *pH dependent selectivity in the phosphoro-Strecker reaction. In the presence of both DAP and ammonia, aminonitrile* ***10*** *is predominantly formed at pH 10, and phosphorylated derivative* ***114*** *at neutral pH.*

The phosphoro-Strecker reaction was observed to exhibit further selectivity: in the direct competition between aldehydes **8** and ketones **242**, the canonical α-H products are predominantly formed, providing a means for the formation of proteinogenic amino acids from complex mixtures of starting materials on the early Earth (***Figure 6.3***).

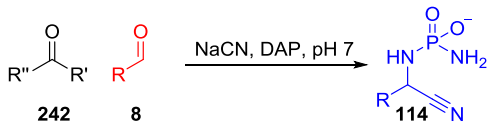


Figure 6.3: *Substrate selectivity in the phosphoro-Strecker reaction: the canonical α-H product* ***114*** *predominantly forms from mixtures of aldehydes* ***8*** *and ketones* ***242****.*

The phosphorylated aminonitrile **114** can be selectively derivatised to allow access to further reactivity. Incubation in aqueous acid affords the corresponding aminonitrile **10**, whilst incubation in alkaline solution provides the phosphorylated amino acid **82**. Reaction of **114** with hydrogen sulfide provides thiolysis product **131** that may be capable of further reaction (and is currently under investigation in the Powner lab). The

synthesis of phosphorylated aminonitriles and their subsequent derivatisations provide an insight into prebiotic routes to activated species that could be used in prebiotic peptide formation.

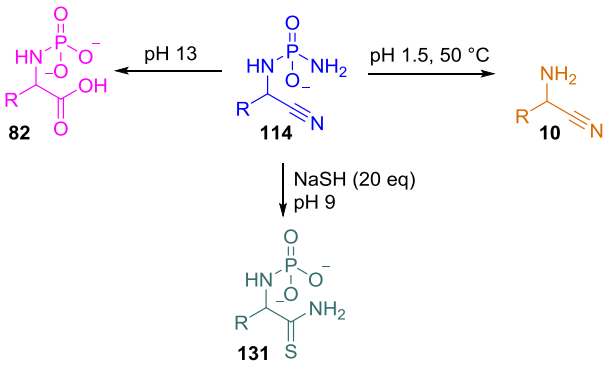


Figure 6.4: *Reactivity of N-phosphoramidate aminonitriles* ***114****, hydrolysing to give aminonitrile* ***10*** *at acidic pH and* ***82*** *at basic pH, and thiolysis to* ***131.***

DAP has previously been implicated as a phosphorylating agent for simple aldehydes and sugars,[153] allowing entry into the triose glycolysis pathway,[154] and for nucleotides, lipids and amino acids (***Figure 6.5***).[208] DAP provides a robust mechanism for the α-phosphorylation of simple aldehydes, providing a point of divergence between triose glycolysis and nucleotide synthesis,[154] which could potentially take place in the same environment as the phosphorylation of other classes of (pre-)biologically relevant molecules.[208] The use of DAP in effecting a number of transformations in different areas of prebiotic chemistry links these various reactions through a common reagent and suggests that further investigations of the reactivity of DAP may prove fruitful in prebiotic systems chemistry.

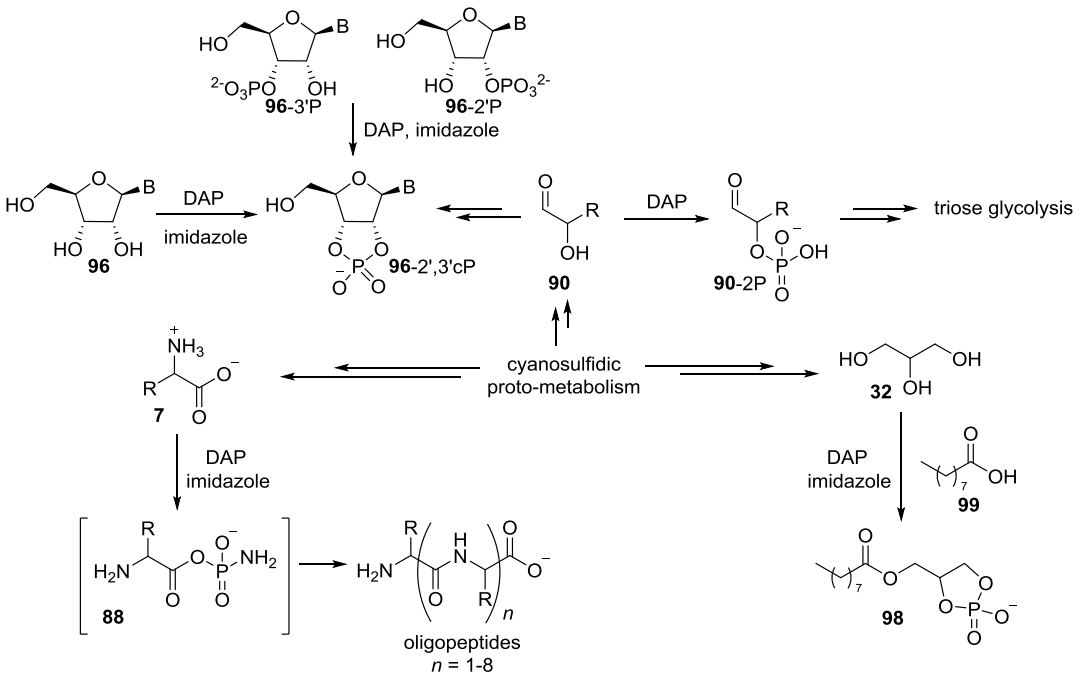


***Figure 6.5**: DAP can be used as a phosphorylating agent across several classes of biologically relevant molecules. Phosphorylation of α-hydroxy aldehydes **90** with DAP allows entry to the triose glycolysis pathway. Nucleoside **96** and nucleotides **96**-3'P and **96**-2'P yield activated 2',3'-cyclic phosphate nucleotides **96**-2',3'cP. The reaction of glycerol **32**, nonanoic acid **99** and DAP provides a cyclophospholipid **98**. Oligopeptides can be detected in the reaction of amino acid **7** with DAP, with **88** as the presumed intermediate.*

The potential prebiotically relevant systems chemistry of DAP indicates that a source of DAP on the early Earth may have been beneficial to the origin of life – but a suitable phosphorus source to allow the formation of DAP must have been available. The prebiotic availability of phosphate has been debated, but the work presented here once again underlines the essential and irreplaceable role of phosphate.[329] One of the aims of prebiotic research is to demonstrate why certain elements were not selected for a biochemical role in early scenarios.[400] The role of phosphate in dictating the structure and function of nucleic acids[233,235] and the localisation and compartmentalisation of metabolites[176] is well-known, but we have also shown here that replacing phosphate with borate does not permit the key rearrangement step in an existing pyrimidine synthesis (***Figure 6.6***).[138] Phosphate fulfils several essential roles in this synthetic route,[138] further emphasising the necessity of a prebiotic source of phosphate.

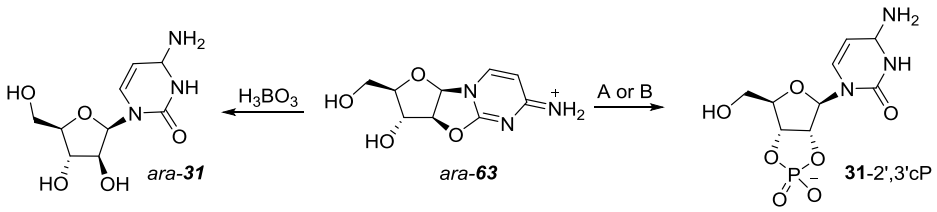


Figure 6.6: *Reported synthesis of cytidine **31**-2′,3′cP by urea-mediated phosphorylation of anhydrocytidine ara-**63** (A: PPi, urea, heat; B: Pi, urea, formamide solution, heat), and formation of non-canonical arabinocytosine ara-**31** from ara-**63** in the presence of H_3BO_3.*

An understanding of what does not work not only provides us with an insight into plausible chemical scenarios on the early Earth, but also constrains future explorations so that they can be more focused and, hopefully, successful.[401] Our attempts presented here to effect the glycosylation of nucleobases using acetylated sugars were unsuccessful, but place further weight on the investigation of other routes to nucleotides, such those that involve the initial formation of a sugar-nucleobase hybrid scaffold.[138,342] A route to canonical purine nucleosides in this style has not yet been completed, although significant advances have been made resulting in the formation of 8-oxo-purine nucleotides.[342] Our attempts to carry out a chemospecific reduction of 5′,8-anhydro purines, however, was unsuccessful. Nevertheless, a route to the canonical purine nucleotides may yet be found, informed by our investigations presented here.

A systems chemistry approach to the synthesis of prebiotically relevant molecules, including a route to amino acids and peptides using the phosphoro-Strecker reaction, emphasises the importance of different methods of phosphorylation, and different sources of phosphate, on the early Earth. The selective installation of nucleophilic phosphate in solid state conditions provides an activated nucleotide 2′,3′-cyclic phosphate with correct sugar stereochemistry caused by reaction and inversion at C-2′ (shown in ***Figure 6.6***);[138] phosphorylation of simple aldehydes with amidophosphates, taking place in aqueous neutral or acidic solution, is an essential step in the prebiotic formation of triose glycolysis metabolites (shown in ***Figure 6.5***);[154] whilst the work presented here links the chemistry of amidophosphates in neutral solution to amino acid synthesis. The conditions and methods used in this nucleoside phosphorylation and reactions with DAP are obviously different: one taking place in the dry state, using urea and ammonium orthophosphate; the other in aqueous conditions. Nevertheless, these methods can be linked.

Two of the most prebiotically plausible routes to orthophosphate – oxidation of schreibersite-derived minerals[2,200] and hydrolysis of volcanic P_4O_{10}[187] – also produce trimetaphosphate, a precursor, *via* ammonolysis,[207] of DAP. Ammonia is thus also a common reagent: in addition to production of DAP, it is also required in the drying phase of urea-mediated phosphorylation of nucleosides, where its volatility drives acidification.[138] In fact, cyclic trimetaphosphate is also readily formed under these conditions: it is a product of the solid state heating of urea and ammonium orthophosphate,[201] thus providing a further link between the synthesis of nucleotides, central (energy) metabolism, and amino acids (***Figure 6.7***).

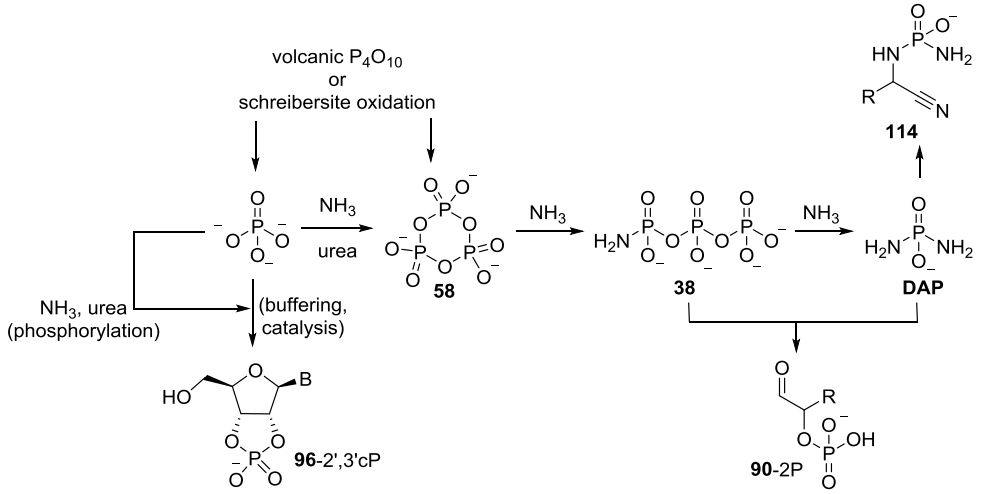


Figure 6.7: *The intergenerational relationships of various prebiotically plausible phosphates and their use in the synthesis of biologically relevant molecules and precursors.*

The fact that overall levels of DAP produced on the early Earth by these methods were likely to be low and localised may have been advantageous. A proposed cyanosulfidic reaction network produces a range of prebiotically relevant molecules,[64] including cyanohydrins which could be used as substrates in the phosphoro-Strecker reaction. With DAP used as a limiting reagent, and so comparatively low conversion to phosphorylated aminonitriles, the majority of species generated in this network are available for transformation into nucleotide, lipid and metabolite precursors. Following a systems chemistry approach, these could have been phosphorylated (if necessary) using a range of different phosphate sources and phosphorylation conditions, to access the complete set of species required for the advent of life.

Linking results obtained during research into the synthesis of the first possible biomolecules with relevant results from other aspects of prebiotic research informs our

understanding of the process that may contributed towards the establishment of life, thereby generating an overview of the chemistry that may have taken place at the origin of life. Connecting different classes of metabolites through the use of common precursors, intermediates and reagents allows a closer understanding of the routes that life may have taken during the first stages of evolution. The work here described is therefore a step towards understanding the processes that governed the chemical evolution of life.

7. Experimental

7.1 General

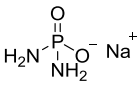
Reactions requiring anhydrous conditions were carried out under an argon atmosphere using oven-dried glassware. Reagents and solcents obtained from commercial sources were used without further purification. Deionised water was obtained from an Elga Option 3 purification system. A Mettler Toledo Seven Compact pH meter with a Mettler Toledo InLab semi-micro pH probe was used to measure solution pH values. Readings for D_2O solutions reported as pD; readings for H_2O and H_2O/D_2O solutions are reported uncorrected. Automated flash column chromatography was carried out using a Biotage Isolera Four purification system and either a Telos Silica Flash-LL column or a Biotage KP-C18-HS Snap Cartridge.

^{1}H NMR spectra were recorded using an internal deuterium lock for residual protons in $CDCl_3$ (δ_H 7.26 ppm) at ambient probe temperature on the following machines: Avance 300 (300 MHz), Avance III 400 (400 MHz), Avance III 600 Cryo (600 MHz) and Avance Neo 700 (700 MHz). Data are presented in the order: chemical shift, integration, multiplicity (s, singlet; d, doublet; dd, doublet of doublets; dt, doublet of triplets; t, triplet; td, triplet of doublets; tt, triplet of triplets; q, quartet; quin, quintet; m, multiplet; ABX, germinal (AB) spin system coupled to one other nucleus (X); ABXY, germinal spin system coupled to two other nuclei), coupling constant, and proton assignment. Assignments were determined either on basis of unambiguous chemical shift or coupling pattern, or by analogy to fully interpreted spectra or related compounds. ^{13}C NMR spectra were recorded using an internal C-13 lock for $CDCl_3$ (δ_C 77.0 ppm) at ambient probe temperature by broadband proton spin decoupling on an Avance III 600 Cryo (150 MHz) machine. ^{31}P NMR spectra were recorded on the following machines: Avance III 400 (162 MHz) and Avance Neo 700 (284 MHz). Chemical shifts are given in ppm on a δ scale relative to δ_{TMS} = 0 ppm. Coupling constants (*J*) are given in Hz to the nearest 0.1 Hz. Infrared spectra (IR) were recorded on a Perkin Elmer Spectrum 100 FT-IR spectrometer.

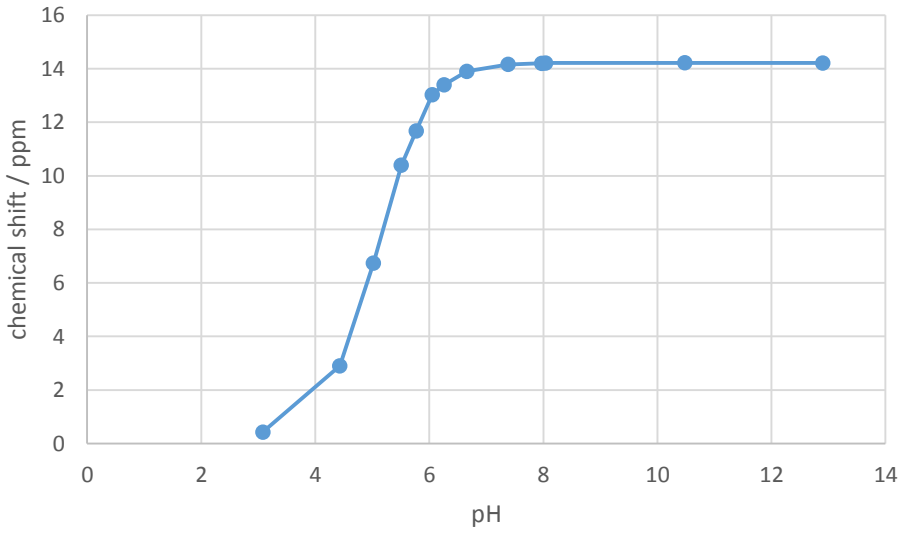
7.2 Experimental section for the phosphoro-Strecker reaction

7.2.1 Synthesis of amidophosphates

Diamidophosphate, DAP, sodium salt:

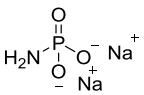


Phenylphosphorodiamidate (10 g, 0.06 mol) was dissolved in sodium hydroxide (4M, 30 mL) and boiled for 20 min. The solution was cooled to room temperature, then added dropwise to rapidly stirring ice-cold ethanol (200 mL). The precipitate was collected by filtration and desiccated over P_2O_5 to give a white solid (6.8 g, 70%). The amount of DAP per unit mass of material was established by relative integration of ^{31}P NMR spectra against a known quantity of inorganic phosphate (as various levels of hydration of DAP could be obtained depending on the extent of drying). ^{31}P NMR (161 MHz, D_2O) δ_P 14.20. HRMS ($[H_4N_2O_2P+2H]^+$) calcd. 97.0106, found 97.0107.



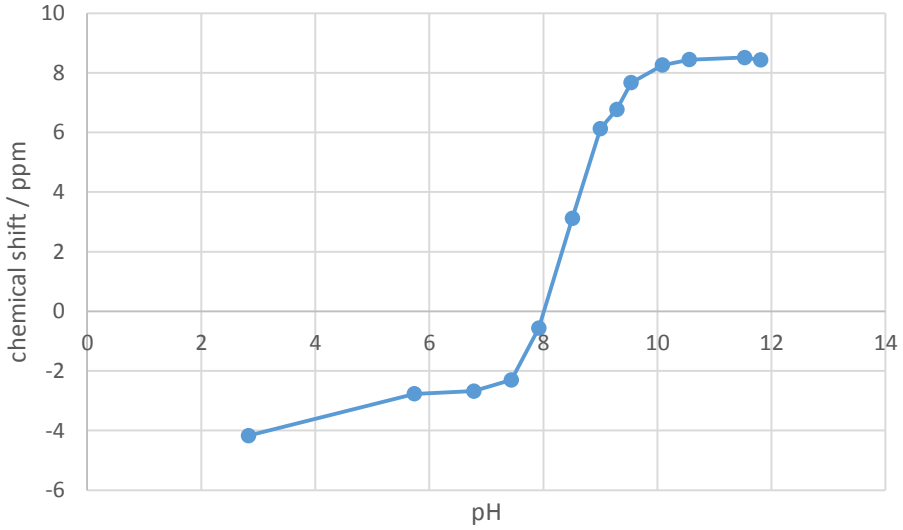
Graph 7.2.1: *a plot of the chemical shift in the ^{31}P NMR spectrum of DAP against pH. A complex mixture of products (from hydrolysis and condensation) were observed at pH <5, which meant chemical shifts at a full range of pH values could not be obtained.*

Monoamidophosphate, MAP, sodium salt:



Dihenylphosphoroamidate (1.7 g, 6.9 mmol) was dissolved in sodium hydroxide (5M, 7 mL) and boiled for 20 min. The solution was cooled to room temperature, then added dropwise to rapidly stirring ice-cold ethanol (150 mL). The precipitate was collected by

filtration and desiccated over P_2O_5 to give a white solid (981.27 mg, 55%). The amount of MAP per unit mass of material was established by relative integration of ^{31}P NMR spectra against a known quantity of inorganic phosphate (as various levels of hydration of MAP could be obtained depending on the extent of drying). ^{31}P NMR (161 MHz, D_2O) δ_P 8.15. m/z (ES-) 96.94 (H_4NO_3P).



Graph 7.2.2: *a plot of the chemical shift in the ^{31}P NMR spectrum of MAP against pH. A complex mixture of products (from hydrolysis and condensation) were observed at pH <8, which meant chemical shifts at a full range of pH values could not be obtained.*

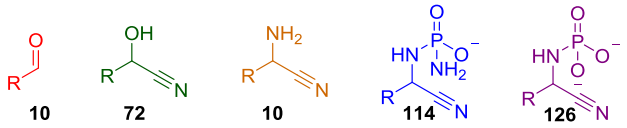
7.2.2 General procedures for investigating the phosphoro-Strecker reaction

General procedure 2.1: Aldehyde (1 eq) was dissolved in water and pentaerythritol standard and NaCN (1.2 eq) added, and adjusted to pH 9.2. DAP solution (4 eq) was added and adjusted to pH 7.0. The pH was monitored over the course of the reaction and periodically adjusted back to pH 7. 500 µL aliquots were periodically removed, freeze-dried, resuspended in D_2O and NMR spectra collected.

General procedure 2.2: Aldehyde (1 eq) was dissolved in water and pentaerythritol standard and NaCN (1.2 eq) added, and adjusted to pH 9.2. DAP solution (4 eq) was added and adjusted to pH 7.0. The pH was monitored over the course of the reaction and periodically adjusted back to pH 7. 500 µL aliquots were periodically removed, diluted with 100 µL D_2O, and NMR spectra collected.

7.2.3 Exploring the optimal reaction conditions

7.2.3.1 Exploring the optimal reaction conditions: pH



*Propanal **8d** ($R=CH_2CH_3$)*

Tables 7.2.1-6 show yields of various species observed by ^{1}H and ^{31}P NMR analysis in the phosphoro-Strecker reaction of propanal **8d** at pH 5-10. These results are summarised in ***Graph 2.1***; representative ^{1}H NMR spectra are shown in ***Figure 2.16***; and representative ^{31}P NMR spectra are shown in ***Figure 2.17***.

NMR data for species observed:

72d: ^{1}H NMR (700 MHz,D_2O) δ_H 4.59 (1H, t, *J* = 6.6 Hz, (C2)-H), 1.84 (2H, app quin, *J* = 7.3 Hz, (C3)-H_2, diagnostic peak), 1.01 (3H, t, *J* = 7.5 Hz, (C4)-H_3).

114d: ^{1}H NMR (700 MHz,D_2O) δ_H 3.98 (1H, app q, *J* = 8.5 Hz, (C2)-H), 1.81-1.77 (2H, m, (C3)-H_2, diagnostic peak), 1.02 (3H, t, *J* = 7.5 Hz, (C4)-H_3). ^{31}P NMR (264 MHz,D_2O) δ_P 10.68.

10d: ^{1}H NMR (700 MHz,D_2O) δ_H 3.86 (1H, dd, *J* = 7.7, 6.1 Hz, (C2)-H), 1.83-1.77 (2H, m, (C3)-H_2, overlaps with **114d**), 1.04-0.99 (3H, m, (C4)-H_3, overlaps with **72d** and **114d**).

t(d)	8d	72d	114d + 10d*	114d (by ^{31}P)	% aldehyde observed	
					sum of known peaks	by integration of methyl peaks
0	-	57%	0%	1%	57%	84%
1	-	35%	17%	17%	52%	85%
2	-	22%	28%	25%	60%	86%
3	-	13%	34%	33%	47%	87%
4	-	11%	38%	35%	49%	87%

Table 7.2.1: *reaction of propanal **8d** (200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature at pH 10, general procedure 2.2. * peaks overlap*

t(d)	8d	72d	114d + 10d*	114d (by ^{31}P)	% aldehyde observed	
					sum of known peaks	by integration of methyl peaks
0	-	55%	0%	1%	55%	79%
1	-	34%	21%	21%	55%	81%
2	-	21%	37%	34%	58%	86%
3	-	14%	43%	43%	56%	84%
4	-	9%	48%	45%	57%	90%
5	-	7%	50%	44%	57%	96%
6	-	4%	53%	44%	58%	93%

Table 7.2.2: *reaction of propanal **8d** (200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature at pH 9, general procedure 2.2. * peaks overlap*

t(d)	8d	72d	114d + 10d*	114d (by ^{31}P)	% aldehyde observed	
					sum of known peaks	by integration of methyl peaks
0	-	55%	3%	3%	57%	79%
1	-	23%	37%	33%	60%	87%
2	-	11%	50%	49%	61%	89%
3	-	6%	56%	57%	62%	92%
4	-	4%	59%	58%	63%	88%
5	-	0%	61%	58%	61%	89%
6	-	0%	63%	60%	63%	92%

Table 7.2.3: *reaction of propanal* ***8d*** *(200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature at pH 8, general procedure 2.2. * peaks overlap*

t(d)	8d	72d	114d + 10d*	114d (by ^{31}P)	% aldehyde observed	
					sum of known peaks	by integration of methyl peaks
0	-	54%	5%	5%	59%	79%
1	-	12%	51%	50%	63%	86%
2	-	5%	59%	59%	64%	87%
3	-	4%	61%	61%	65%	88%
4	-	3%	60%	63%	63%	84%

Table 7.2.4: *reaction of propanal* ***8d*** *(200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature at pH 7, general procedure 2.2. * peaks overlap*

t(d)	8d	72d	114d + 10d*	114d (by ^{31}P)	% aldehyde observed		% DAP remaining (by ^{31}P)
					sum of known peaks	by integration of methyl peaks	
0	-	62%	2%	2%	65%	85%	95%
1	-	22%	48%	35%	70%	93%	71%
2	-	12%	60%	43%	71%	88%	60%
3	-	11%	61%	50%	71%	94%	54%
4	-	10%	61%	55%	71%	93%	43%
6	-	10%	63%	48%	73%	91%	26%

Table 7.2.5: *reaction of propanal* ***8d****(200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature at pH 6, general procedure 2.2.The reaction mixture started to precipitate after 3 days, albeit to a lesser extent than the analogous reaction at pH 5; analysis of the precipitate revealed it contained the same species as the analogous precipitate at pH 5.*

** peaks overlap*

t(d)	8d	72d	114d + 10d*	114d (by ^{31}P)	% aldehyde observed		% DAP remaining (by ^{31}P)
					sum of known peaks	by integration of methyl peaks	
0	-	67%	0%	0%	67%	72%	89%
1	-	41%	18%	16%	59%	72%	37%
2	-	42%	30%	24%	72%	91%	19%
3	-	37%	32%	22%	69%	88%	4%
4	-	31%	36%	23%	67%	89%	0%
6	-	30%	34%	23%	64%	83%	0%
8	-	34%	35%	19%	70%	94%	0%

Table 7.2.6: *reaction of propanal* ***8d*** *(200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature at pH 5, general procedure 2.2. The reaction mixture started to precipitate after 3 days. Analysis of the precipitate revealed it was made up of the same species as the reaction at pH 6. * peaks overlap*

Analysis of the precipitate formed (using the phosphoro-Strecker reaction of propanal at pH 6 as a representative example): The 3 day NMR sample of the phosphoro-Strecker reaction of propanal at pH 6 was transferred to an Eppendorf vial and centrifuged. The supernatant was carefully pipetted off and the precipitate dissolved in D_2O (500 µL), and adjusted from pH 6.0 to 12.3 to solubilise the solid. The resulting solution was then analysed by NMR. The presence of pentaerythritol and phosphorylated products in this NMR demonstrates that calculations of yield based on relative integration to the internal standard in ^{1}H NMR spectra and by relative integration of species seen in ^{31}P NMR spectra are inaccurate, but are nevertheless useful in illustrating a general trend in yields over the course of the reaction. Moreover, the relative proportions of cyanohydrin and **114d** in solution indicate that the conversion of cyanohydrin to phosphordiamidate aminonitrile is slower at pH 6 than at pH 7.

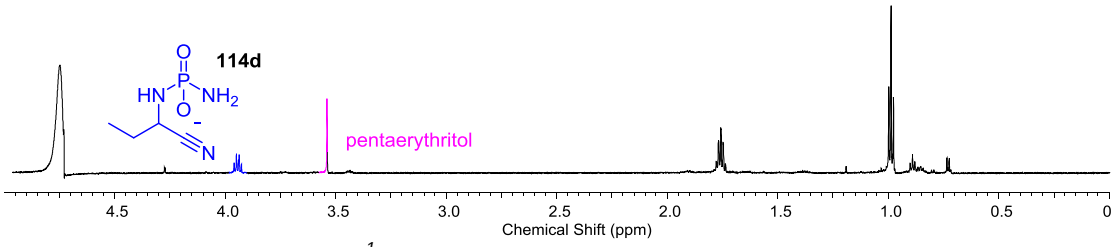


Figure 7.2.1: *Water-supressed ^{1}H NMR spectrum (H_2O:D_2O 9:1, 0.00–5.00 ppm, 700 MHz) to show the composition of the precipitate formed in the reaction of propanal* ***8d*** *(200 mM) with sodium cyanide (1.2 eq.) and DAP (4 eq.) at pH 6.0 at room temperature after 3 d.*

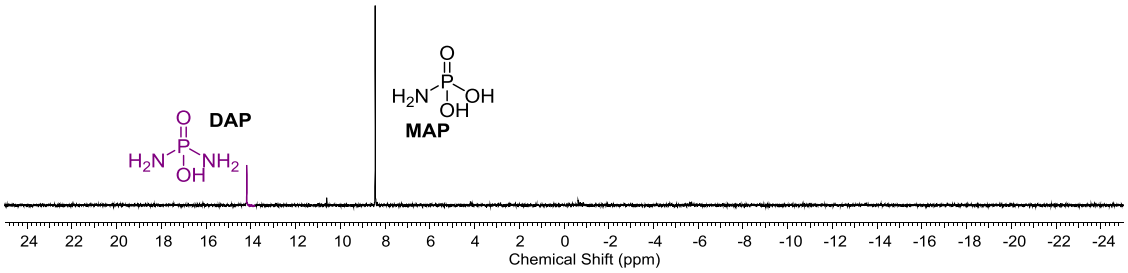


Figure 7.2.2: *^{31}P NMR spectrum (284 MHz, D_2O, -25–25 ppm) to show the composition of the precipitate formed in the reaction of propanal **8d** (200mM) with sodium cyanide (1.2 eq.) and DAP (4 eq.) at pH 6.0 at room temperature after 3 d.*

*Isobutyraldehyde **8e** (R=CH(CH$_3$)$_2$)*

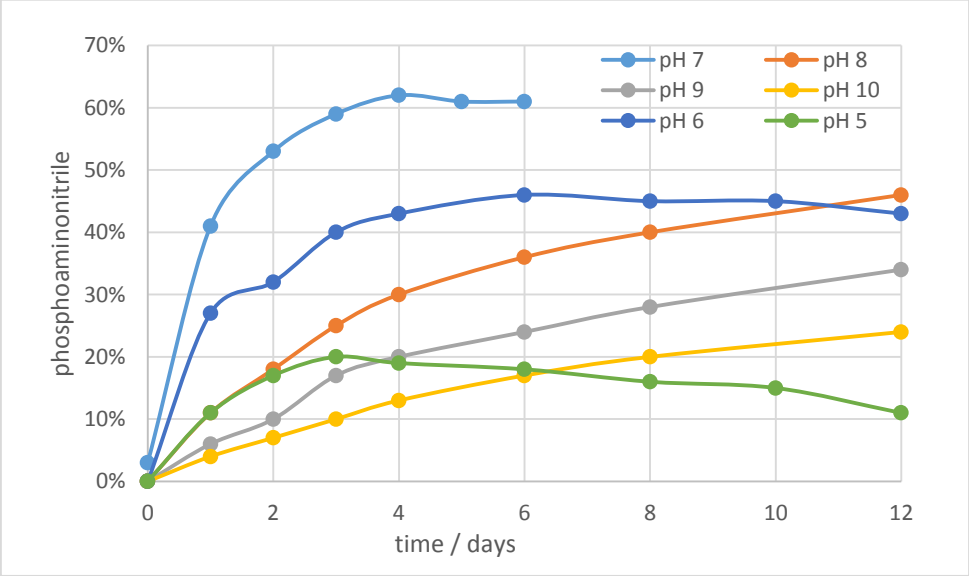
Tables 7.2.7-12 show yields of various species observed by ^{1}H and ^{31}P NMR analysis in the phosphoro-Strecker reaction of isobutyraldehyde **8e** at pH 5-10. These results are summarised in ***Graph 7.2.3***.

Diagnostic NMR data for species observed:

114e: ^{1}H NMR (400 MHz, D_2O) δ_H 3.89 (1H, dd, *J* = 9.3, 6.3 Hz, (C2)-H), 2.00-1.94 (1H, m, (C3)-H, overlaps with **10e**). ^{31}P NMR (162 MHz, D_2O) δ_P 10.88.

10e: ^{1}H NMR (400 MHz, D_2O) δ_H 3.75 (1H,d, *J* = 6.0 Hz, (C2)-H), 2.00-1.94 (1H, m, (C3)-H, overlaps with **114e**).

72e: ^{1}H NMR (400 MHz, D_2O) δ_H 2.07-2.02 (1H, m, (C3)-H).



Graph 7.2.3*: Rate of reaction of isobutyraldehyde **8e** (200 mM) with NaCN (1.2 eq) and then DAP (4.0 eq) at a range of pH values. Yields calculated by relative integration of ^{31}P NMR spectra: this method means that the yields for reaction at pH 5 are inaccurate after 3 days.*

t(d)	8e	72e	114e + 10e*	114e (by ^{31}P)	% aldehyde observed
0	-	44%	0%	0	44%
1	-	38%	8%	4%	46%
2	-	34%	13%	7%	47%
3	-	30%	19%	10%	49%
4	-	26%	23%	13%	49%
6	-	20%	30%	17%	50%
8	-	16%	34%	20%	50%
12	-	11%	36%	24%	47%

Table 7.2.7: *reaction of isobutyraldehyde* ***8e*** *(200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature at pH 10, general procedure 2.2. * peaks overlap*

t(d)	8e	72e	114e + 10e*	114e (by ^{31}P)	% aldehyde observed
0	-	43%	0%	0%	43%
1	-	35%	11%	6%	46%
2	-	31%	16%	10%	47%
3	-	24%	24%	17%	48%
4	-	19%	31%	20%	50%
6	-	14%	35%	24%	49%
8	-	9%	41%	28%	50%
12	-	7%	43%	34%	50%

Table 7.2.8: *reaction of isobutyraldehyde* ***8e*** *(200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature at pH 9, general procedure 2.2. * peaks overlap*

t(d)	8e	72e	114e + 10e*	114e (by ^{31}P)	% aldehyde observed
0	-	42%	0%	0%	42%
1	-	33%	14%	11%	47%
2	-	26%	24%	18%	50%
3	-	17%	32%	25%	49%
4	-	13%	37%	30%	50%
6	-	6%	46%	36%	52%
8	-	5%	47%	40%	52%
12	-	5%	50%	46%	55%

Table 7.2.9: *reaction of isobutyraldehyde* ***8e*** *(200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature at pH 8, general procedure 2.2. * peaks overlap*

t(d)	8e	72e	114e + 10e*	114e (by ^{31}P)	% aldehyde observed
0	-	-	5%	2%	3%
1	-	-	41%	9%	36%
2	-	-	51%	7%	45%
3	-	-	59%	6%	52%
4	-	-	63%	5%	55%
5	-	-	61%	3%	55%
6	-	-	65%	4%	57%

Table 7.2.10: *reaction of isobutyraldehyde* ***8e*** *(200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature at pH 7, general procedure 2.1. * peaks overlap*

t(d)	8e	72e	114e + 10e*	114e (by ^{31}P)	% aldehyde observed	% DAP remaining (by ^{31}P)
0	-	58%	0%	0%	58%	95%
1	-	36%	27%	27%	63%	76%
2	-	22%	40%	32%	62%	60%
3	-	14%	49%	40%	63%	51%
4	-	9%	55%	43%	64%	41%
6	-	7%	60%	46%	67%	31%
8	-	10%	58%	45%	68%	21%
10	-	9%	59%	45%	68%	11%
12	-	11%	60%	43%	71%	4%

Table 7.2.11: *reaction of isobutyraldehyde* ***8e*** *(200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature at pH 6, general procedure 2.2.NB: no precipitation was observed over the course of the reaction at pH 6. * peaks overlap*

t(d)	8e	72e	114e + 10e*	114e (by ^{31}P)	% aldehyde observed	% DAP remaining (by ^{31}P)
0	-	53%	0%	0%	53%	83%
1	-	40%	15%	11%	55%	36%
2	-	33%	21%	17%	54%	18%
3	-	30%	22%	20%	52%	9%
4	-	31%	24%	19%	55%	4%
6	-	29%	26%	18%	55%	0%
8	-	30%	22%	16%	52%	0%
10	-	31%	22%	15%	53%	0%
12	-	29%	21%	11%	50%	0%

Table 7.2.12: *reaction of isobutyraldehyde* ***8e*** *(200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature at pH 5, general procedure 2.2. The reaction mixture started to precipitate after 3 days, and was analysed in an analogous fashion to that of propanal. * peaks overlap*

7.2.3.2 Comparison and competition with the conventional Strecker reaction

Results of competition reactions between DAP and NH_4Cl the phosphoro-Strecker reactions of propanal **8d** and isobutyraldehyde **8e** are summarised in ***Table 2.1***.

Propanal ***8d*** *($R=CH_2CH_3$)*

Representative ^{1}H NMR spectra for the conventional Strecker reaction of popanal **8d** at pH 10 and pH 7 (yields shown in ***Table 7.2.13*** and ***Table 7.2.14***) are shown in ***Figure 2.18***.

t(d)	8d	72d	10d	% aldehyde observed	
				sum of known peaks	by integration of methyl peaks
0	-	44%	13%	57%	84%
1	-	5%	48%	53%	74%
2	-	6%	48%	54%	82%
3	-	6%	48%	54%	84%
4	-	7%	46%	53%	81%
5	-	7%	45%	52%	80%
6	-	7%	45%	52%	83%

Table 7.2.13: *reaction of propanal* ***8d*** *(200 mM) with NaCN (1.2 eq) and NH_4Cl (4 eq) in H_2O at room temperature at pH 10, general procedure 2.2.*

t(d)	8d	72d	10d	% aldehyde observed	
				sum of known peaks	by integration of methyl peaks
0	-	55%	0%	55%	81%
1	-	51%	6%	57%	85%
2	-	51%	8%	59%	88%
3	-	49%	8%	57%	89%
4	-	50%	10%	60%	90%
5	-	50%	13%	63%	90%
6	-	46%	13%	59%	87%

Table 7.2.14: *reaction of propanal* ***8d*** *(200 mM) with NaCN (1.2 eq) and NH_4Cl (4 eq) in H_2O at room temperature at pH 7, general procedure 2.2.*

Yields shown in ***Table 7.2.15*** and ***Table 7.2.16*** are summarised in ***Table 2.1***.

t(d)	8d	72d	114d + 10d*	114d (by ^{31}P)	% aldehyde observed	
					sum of known peaks	by integration of methyl peaks
0	-	45%	<1%	13%	58%	82%
1	-	4%	3%	47%	51%	84%
2	-	4%	5%	50%	54%	82%
4	-	3%	7%	49%	52%	82%
5	-	4%	7%	49%	53%	83%
6	-	4%	8%	47%	51%	82%

Table 7.2.15: *competition reaction of propanal* ***8d*** *(200 mM) with NaCN (1.2 eq), DAP (4 eq) and NH_4Cl (4 eq) in H_2O at room temperature at pH 10, general procedure 2.2. * peaks overlap*

t(d)	8d	72d	114d + 10d*	114d (by ^{31}P)	% aldehyde observed	
					sum of known peaks	by integration of methyl peaks
0	-	54%	3%	4%	58%	84%
1	-	11%	38%	46%	57%	92%
2	-	8%	43%	53%	61%	96%
3	-	8%	46%	56%	64%	98%
4	-	6%	49%	56%	62%	95%
5	-	5%	50%	56%	61%	94%

Table 7.2.16: *competition reaction of propanal* ***8d*** *(200 mM) with NaCN (1.2 eq), DAP (4 eq) and NH_4Cl (4 eq) in H_2O at room temperature at pH 7, general procedure 2.2. * peaks overlap*

*Isobutyraldehyde **8e** (R=CH(CH$_3$)$_2$)*

t(d)	8e	72e	10e	% aldehyde observed	
				sum of known peaks	by integration of methyl peaks
0	-	38%	4%	42%	51%
1	-	9%	36%	45%	60%
2	-	6%	42%	48%	61%
3	-	5%	44%	49%	61%
4	-	4%	43%	47%	61%
5	-	4%	42%	46%	61%
6	-	6%	43%	49%	64%

***Table 7.2.17:** reaction of isobutyraldehyde **8e** (200 mM) with NaCN (1.2 eq) and NH$_4$Cl (4 eq) in H$_2$O at room temperature at pH 10, general procedure 2.2.*

t(d)	8e	72e	10e	% aldehyde observed	
				sum of known peaks	by integration of methyl peaks
0	-	42%	0%	42%	53%
1	-	39%	5%	44%	55%
2	-	37%	5%	42%	54%
3	-	35%	6%	41%	52%
4	-	30%	6%	36%	49%
6	-	28%	6%	34%	44%

***Table 7.2.18:** reaction of isobutyraldehyde **8e** (200 mM) with NaCN (1.2 eq) and NH$_4$Cl (4 eq) in H$_2$O at room temperature at pH 7, general procedure 2.2.*

t(d)	8e	72e	114e + 10e*	114e (by ^{31}P)	% aldehyde observed	
					sum of known peaks	by integration of methyl peaks
0	-	42%	0%	6%	48%	56%
1	-	7%	2%	41%	48%	57%
2	-	6%	3%	45%	51%	57%
3	-	5%	3%	46%	51%	62%
4	-	5%	4%	46%	51%	61%
6	-	5%	5%	44%	49%	62%
8	-	6%	6%	44%	50%	61%
12	-	6%	7%	42%	48%	61%
14	-	6%	8%	42%	48%	60%

***Table 7.2.19:** competition reaction of isobutyraldehyde **8e** (200 mM) with NaCN (1.2 eq), DAP (4 eq) and NH$_4$Cl (4 eq) in H$_2$O at room temperature at pH 10, general procedure 2.2. * peaks overlap*

t(d)	8e	72e	114e + 10e*	114e (by ^{31}P)	% aldehyde observed	
					sum of known peaks	by integration of methyl peaks
0	-	41%	2%	3%	44%	55%
1	-	14%	26%	37%	51%	57%
2	-	8%	33%	44%	52%	58%
3	-	6%	36%	47%	53%	58%
4	-	6%	38%	47%	53%	56%
6	-	4%	39%	49%	53%	60%
8	-	4%	41%	51%	55%	61%

Table 7.2.20: *competition reaction of isobutyraldehyde* ***8e*** *(200 mM) with NaCN (1.2 eq), DAP (4 eq) and NH_4Cl (4 eq) in H_2O at room temperature at pH 7, general procedure 2.2.* peaks overlap*

7.2.3.3 Exploring the optimal reaction conditions: concentration

Propanal ***8e*** *($R=CH_2CH_3$)*

Table 7.2.4 shows reaction of propanal **8d** (200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) at pH 7. The yields shown in ***Table 7.2.4***, ***Table 7.2.22*** and ***Table 7.2.23*** are summarised in ***Graph 2.2***.

t(d)	8d	72d	114d + 10d*	114d (by ^{31}P)	% aldehyde observed	
					sum of known peaks	by integration of methyl peaks
0	-	54%	2%	3%	56%	80%
1	-	25%	34%	39%	59%	86%
2	-	15%	46%	53%	61%	85%
3	-	8%	54%	59%	62%	87%
4	-	6%	55%	59%	61%	88%
5	-	5%	57%	62%	62%	88%
6	-	5%	58%	63%	63%	88%
7	-	6%	60%	64%	66%	91%

Table 7.2.21: *reaction of propanal* ***8d*** *(100 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature, general procedure 2.2, attempt 1. * peaks overlap*

t(d)	8d	72d	114d + 10d*	114d (by ^{31}P)	% aldehyde observed	
					sum of known peaks	by integration of methyl peaks
0	-	53%	3%	3%	56%	81%
1	-	23%	37%	38%	60%	85%
2	-	12%	51%	50%	63%	90%
3	-	9%	54%	55%	63%	86%
4	-	5%	57%	56%	62%	90%
5	-	6%	57%	59%	63%	89%
7	-	6%	59%	61%	65%	87%

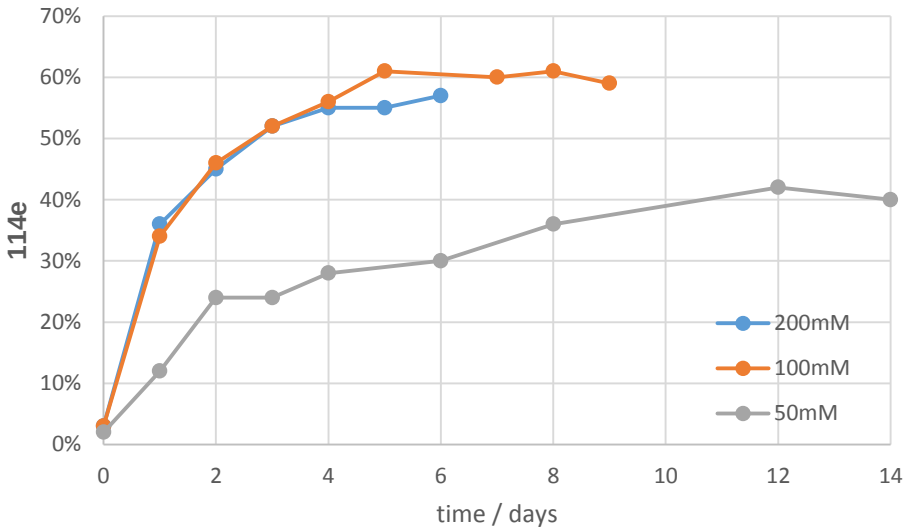
Table 7.2.22: *reaction of propanal* ***8d*** *(100 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature, general procedure 2.2, attempt 2. * peaks overlap*

t(d)	8d	72d	114d + 10d*	114d (by ^{31}P)	% aldehyde observed	
					sum of known peaks	by integration of methyl peaks
0	-	54%	0%	2%	54%	80%
1	-	31%	26%	27%	57%	85%
2	-	23%	40%	41%	63%	88%
3	-	17%	46%	47%	63%	88%
4	-	12%	51%	51%	63%	90%
5	-	10%	53%	54%	63%	90%
6	-	10%	58%	56%	68%	95%
7	-	7%	58%	58%	65%	93%

Table 7.2.23: *reaction of propanal* ***8d*** *(50 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature, general procedure 2.2. * peaks overlap*

Isobutyraldehyde ***8e*** *(R=CH(CH$_3$)$_2$)*

Table 7.2.10 shows reaction of isobutyraldehyde **8e** (200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) at pH 7. The yields shown in ***Table 7.2.10***, ***Table 7.2.24*** and ***Table 7.2.25*** are summarised in ***Graph 7.2.4***.



Graph 7.2.4*: Rate of reaction of isobutyraldehyde* ***8e*** *(50-200 mM) with NaCN (1.2 eq) and then DAP (4.0 eq) at pH 7. Yields calculated by relative integration of ^{31}P NMR spectra.*

t(d)	**114e** (by ^{31}P)
0	3%
1	34%
2	46%
3	52%
4	56%
5	61%
7	60%

Table 7.2.24: *reaction of isobutyraldehyde* ***8e*** *(100 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature, general procedure 2.1. The ^{1}H NMR spectra for this reaction all display broadened peaks (possibly due to metal contamination) and cannot be accurately integrated; the ^{31}P spectra appear normal.*

t(d)	**8e**	**72e**	**114e** + **10e***	**114e** (by ^{31}P)	% aldehyde observed
0	-	41%	0%	2%	41%
1	-	33%	12%	12%	45%
2	-	23%	22%	24%	45%
3	-	17%	30%	24%	47%
4	-	13%	34%	28%	47%
6	-	10%	39%	30%	49%
8	-	7%	42%	36%	49%
12	-	6%	47%	42%	53%
14	-	5%	47%	40%	52%

Table 7.2.25: *reaction of isobutyraldehyde* ***8e*** *(50 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature, general procedure 2.2. * peaks overlap*

7.2.3.4 Exploring the optimal reaction conditions: DAP stoichiometry

Propanal ***8e*** *(R=CH_2CH_3)*

Table 7.2.4 shows reaction of propanal **8d** (200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) at pH 7. The yields shown in ***Table 7.2.4***, ***Table 7.2.26*** and ***Table 7.2.27*** are summarised in ***Graph 2.3***.

t(d)	**8d**	**72d**	**114d** + **10d***	**114d** (by ^{31}P)	% aldehyde observed	
					sum of known peaks	by integration of methyl peaks
0	-	55%	3%	3%	58%	77%
1	-	25%	36%	30%	61%	81%
2	-	15%	49%	39%	63%	84%
3	-	7%	54%	43%	61%	85%
4	-	6%	56%	45%	63%	84%
5	-	6%	57%	46%	63%	88%
6	-	7%	56%	46%	62%	84%

Table 7.2.26: *reaction of propanal* ***8d*** *(200 mM) with NaCN (1.2 eq) and DAP (2.0 eq) in H_2O at room temperature at pH 7, general procedure 2.2. * peaks overlap*

t(d)	8d	72d	114d + 10d*	114d (by ^{31}P)	% aldehyde observed	
					sum of known peaks	by integration of methyl peaks
0	-	55%	2%	2%	57%	77%
1	-	31%	28%	20%	59%	83%
2	-	21%	42%	29%	63%	85%
3	-	13%	48%	33%	61%	86%
4	-	11%	52%	35%	63%	83%
5	-	6%	55%	36%	61%	86%
6	-	6%	56%	37%	62%	85%

Table 7.2.27: *reaction of propanal* ***8d*** *(200 mM) with NaCN (1.2 eq) and DAP (1.2 eq) in H_2O at room temperature at pH 7, general procedure 2.2. * peaks overlap*

Isobutyraldehyde ***8e*** *(R=CH(CH$_3$)$_2$)*

Table 7.2.10 shows reaction of isobutyraldehyde **8e** (200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) at pH 7.

t(d)	8e	72e	114e + 10e*	114e (by ^{31}P)	% aldehyde observed
0	-	39%	2%	0%	41%
1	-	31%	14%	11%	45%
2	-	24%	23%	19%	47%
3	-	17%	30%	24%	47%
4	-	13%	38%	29%	51%
6	-	7%	43%	32%	50%
8	-	4%	47%	35%	50%
12	-	4%	49%	36%	53%
14	-	3%	49%	37%	52%

Table 7.2.28: *reaction of isobutyraldehyde* ***8e*** *(200 mM) with NaCN (1.2 eq) and DAP (2.0 eq) in H_2O at room temperature at pH 7, general procedure 2.2. * peaks overlap*

t(d)	8e	72e	114e + 10e*	114e (by ^{31}P)	% aldehyde observed
0	-	40%	0%	0%	40%
1	-	35%	8%	6%	43%
2	-	30%	14%	12%	44%
3	-	26%	20%	16%	46%
4	-	23%	22%	19%	45%
6	-	18%	29%	24%	37%
8	-	12%	36%	28%	48%
12	-	8%	41%	32%	49%

Table 7.2.29: *reaction of isobutyraldehyde* ***8e*** *(200 mM) with NaCN (1.2 eq) and DAP (1.2 eq) in H_2O at room temperature at pH 7, general procedure 2.2. * peaks overlap*

Exploring the optimal reaction conditions: propanal concentration and DAP stoichiometry

t(d)	8d	72d	114d + 10d*	114d (by ^{31}P)	% aldehyde observed	
					sum of known peaks	by integration of methyl peaks
0	-	25%	0%	0%	25%	67%
1	-	19%	8%	7%	27%	79%
2	-	11%	12%	12%	23%	69%
3	-	9%	16%	14%	25%	80%
4	-	7%	18%	14%	25%	76%
5	-	4%	19%	16%	23%	79%

Table 7.2.30: *reaction of propanal (800 mM) with NaCN (1.2 eq) and DAP (1.0 eq) in H_2O at room temperature at pH 7, general procedure 2.2.* ***Figure 7.2.3*** *shows the ^{1}H NMR spectra for this reaction. * peaks overlap*

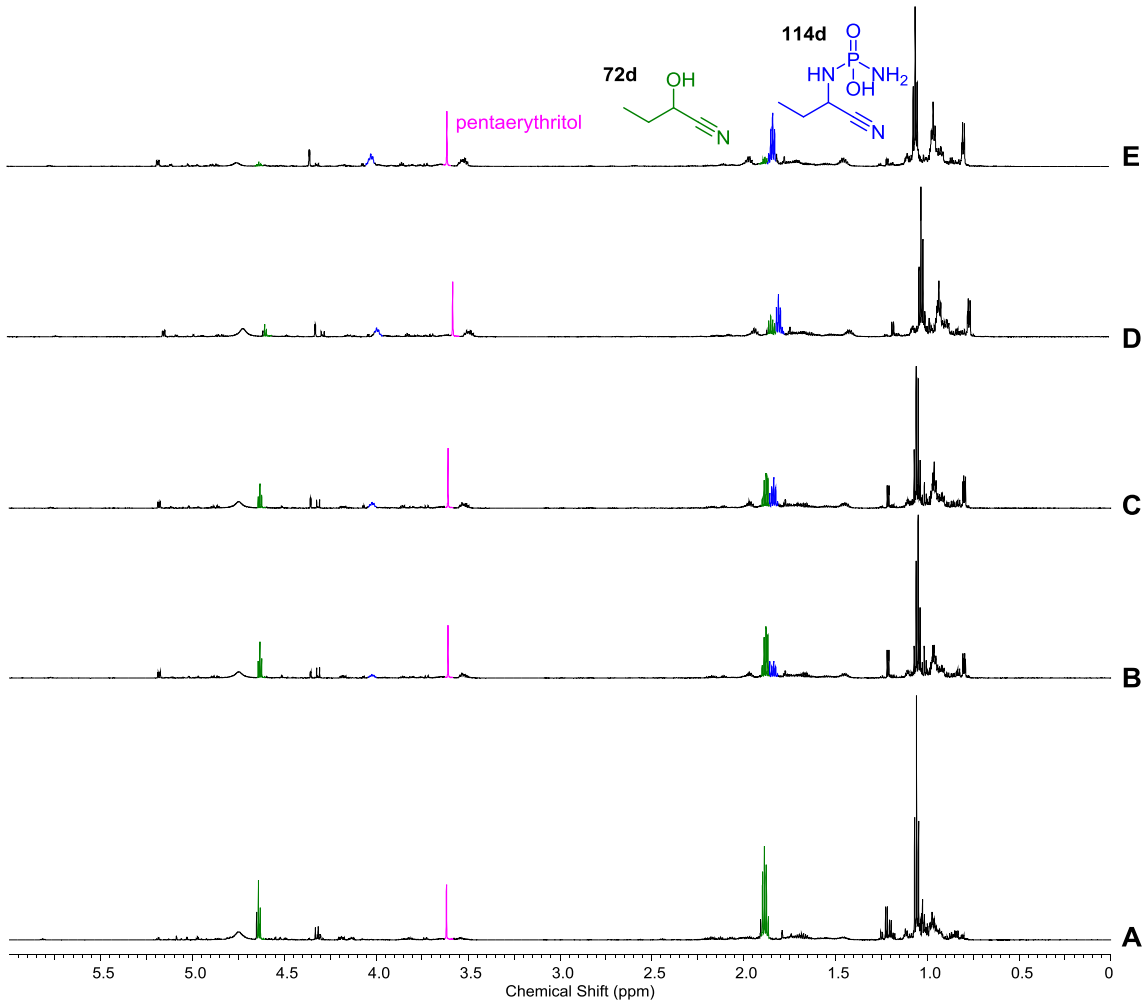


Figure 7.2.3: *Water-supressed ^{1}H NMR spectra (H_2O:D_2O 9:1, 0.00–6.00 ppm, 700 MHz) showing the reaction of propanal* ***8d*** *(800mM) with sodium cyanide (1.2 eq.) and DAP (1 eq.) at pH 7.0 at room temperature after:* ***A****) 30 minutes;* ***B****) 1 d;* ***C****) 2 d;* ***D****) 3 d;* ***D****) 5d.*

t(d)	**72d** (based on aldehyde present)	**114d** (based on aldehyde present)	**114d** (% of potential yield)	**114d** (by ^{31}P)	% aldehyde observed	
					sum of known peaks	by integration of methyl peaks
0	20%	0%	0%	0%	20%	55%
1	14%	3%	6%	4%	17%	49%
2	10%	4%	8%	8%	14%	44%
3	10%	7%	14%	11%	16%	54%
4	8%	8%	16%	13%	16%	51%
5	6%	8%	16%	14%	14%	47%

Table 7.2.31: *reaction of propanal* ***8d*** *(1.6 M) with NaCN (1.2 eq) and DAP (0.5 eq) in H_2O at room temperature at pH 7, general procedure 2.2.* ***Figure 2.19*** *shows the ^{1}H-^{31}P HMBC spectrum for this reaction after 5 d.*

7.2.4 Competition with ketones

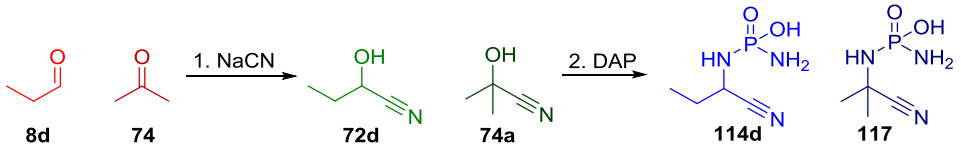
The yields shown in ***Table 7.2.32*** and ***Table 7.2.33*** are summarised in ***Figure 2.20***.

NMR characterisation data for species observed in the phosphoro-Strecker reaction of acetone **74**:

74a: ^{1}H NMR (700 MHz,D_2O) δ_H 1.65 (6H, s).

117: ^{1}H NMR (700 MHz,D_2O) δ_H 1.63 (6H, d, J = 1.63 Hz). ^{31}P NMR (264 MHz,D_2O) δ_P 8.32.

^{1}H NMR spectra shown in ***Figure 7.2.4***; ^{31}P NMR spectra shown in ***Figure 7.2.5***.



t (d)	**74**	**74a**	**117**	**117** (by ^{31}P)	% ketone observed	**72d**	**114d** (by ^{31}P)	**114d** + **10d**	% aldehyde observed
0	44%	20%	0%	2%	64%	58%	4%	5%	64%
1	40%	1%	21%	11%	62%	7%	47%	61%	68%
2	34%	1%	19%	11%	54%	5%	47%	63%	68%
3	32%	1%	18%	10%	51%	4%	48%	64%	68%
4	20%	1%	12%	8%	33%	5%	52%	63%	68%

Table 7.2.32: *competition between propanal* ***8d*** *(200 mM) and acetone* ***74*** *(200 mM) with NaCN (1.2 eq) and DAP (4 eq) in H_2O at room temperature at pH 7, general procedure 2.2.*

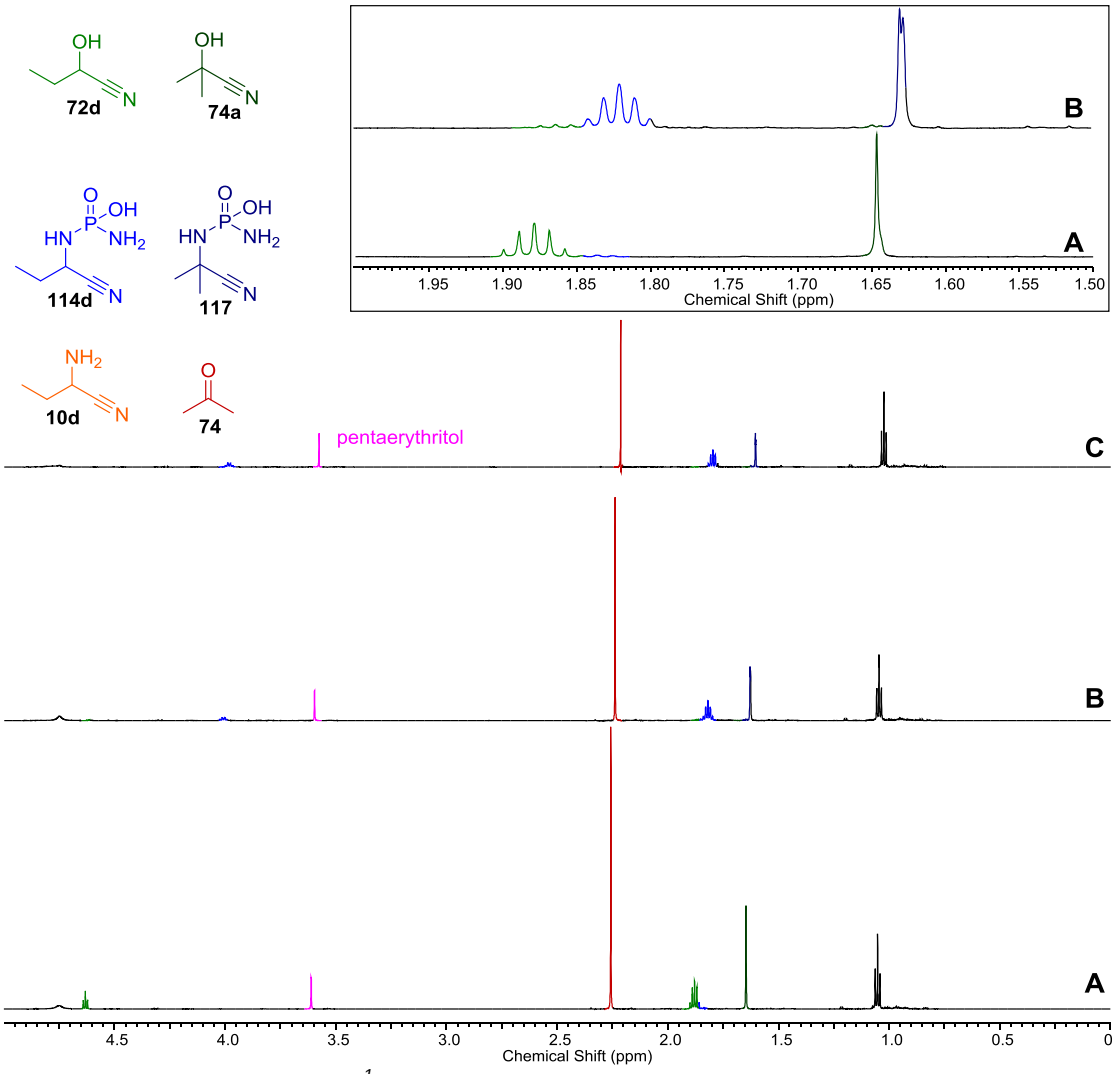


Figure 7.2.4: *Water-supressed ^{1}H NMR spectra (700 MHz; H_2O:D_2O 9:1, 0.00–5.00 ppm; inset 1.5-2.0 ppm) showing the reaction of propanal **8d** (200mM) and acetone **74** (200 mM) with sodium cyanide (1.2 eq.) and DAP (4 eq) at pH 7 at room temperature after: **A**) 30 minutes, **B**) 1 d, and **C**) 4 d.*

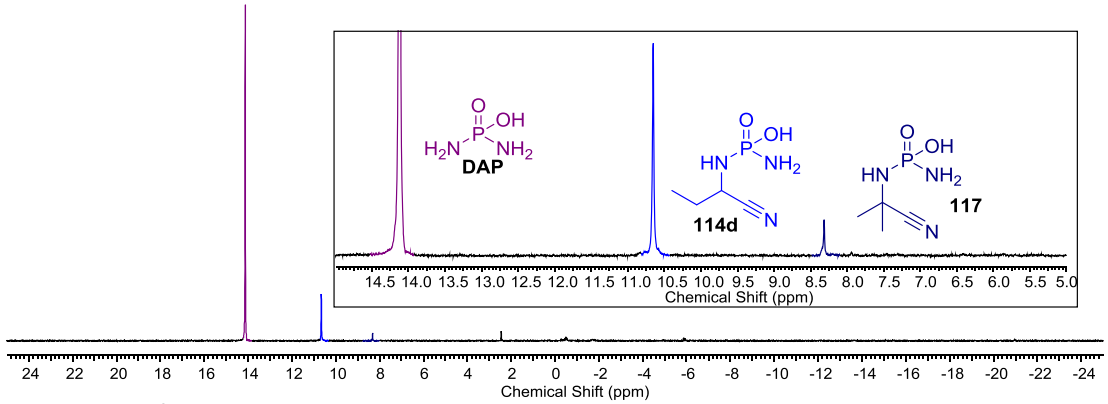


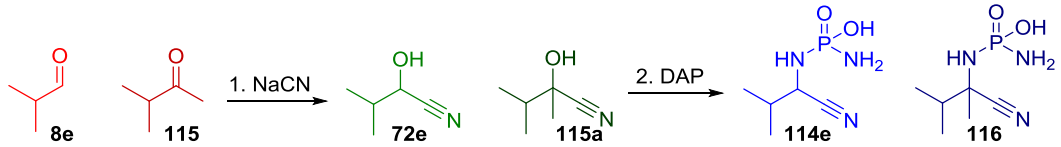
Figure 7.2.5: *^{31}P NMR spectrum (284 MHz, D_2O, -25–25 ppm; inset 5-15 ppm) showing the reaction of propanal **8d** (200mM) and acetone **74** (200 mM) with sodium cyanide (1.2 eq.) and DAP (4 eq.) at pH 7.0 after 4 d at room temperature.*

NMR characterisation data for species observed in the phosphoro-Strecker reaction of 3-methylbutanone **115**:

115a: ^{1}H NMR (700 MHz,D_2O) δ_H 1.58 (3H, s).

116: ^{1}H NMR (700 MHz,D_2O) δ_H 1.56 (3H, s). ^{31}P NMR (264 MHz,D_2O) δ_P 8.61.

^{1}H NMR spectra shown in ***Figure 7.2.7***; ^{31}P NMR spectra shown in ***Figure 7.2.6***.



t(d)	115	115a	116	116 (by ^{31}P)	% Ketone observed	72e	114e (by ^{31}P)	114e + 10e*	% aldehyde observed
0	23	22	0	0%	45	44	1%	20	64
1	25	21	5	4%	52	23	29%	57	81
2	21	18	6	5%	46	15	37%	64	80
3	20	17	7	5%	43	12	40%	70	82
4	17	16	6	4%	40	12	42%	73	85
5	13	9	5	4%	27	8	43%	61	69

Table 7.2.33: *competition reaction of isobutyraldehyde* ***8e*** *(200 mM) and 2-methyl-butanone* ***115*** *(200 mM) with NaCN (1.2 eq) and DAP (4 eq) in H_2O at room temperature at pH 7, general procedure 2.2.peaks overlap*

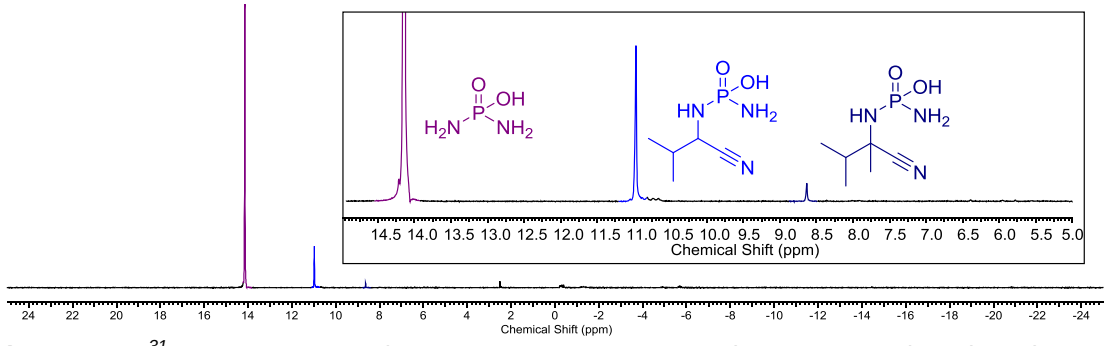


Figure 7.2.6*: ^{31}P NMR spectrum (284 MHz, D_2O, -25–25 ppm; inset 5-15 ppm) to show the reaction of isobutyraldehyde (200 mM) and 3-methyl-butanone (200 mM) with sodium cyanide (1.2 eq.) and DAP (4 eq.) at pH 7.0 after 4 d at room temperature.*

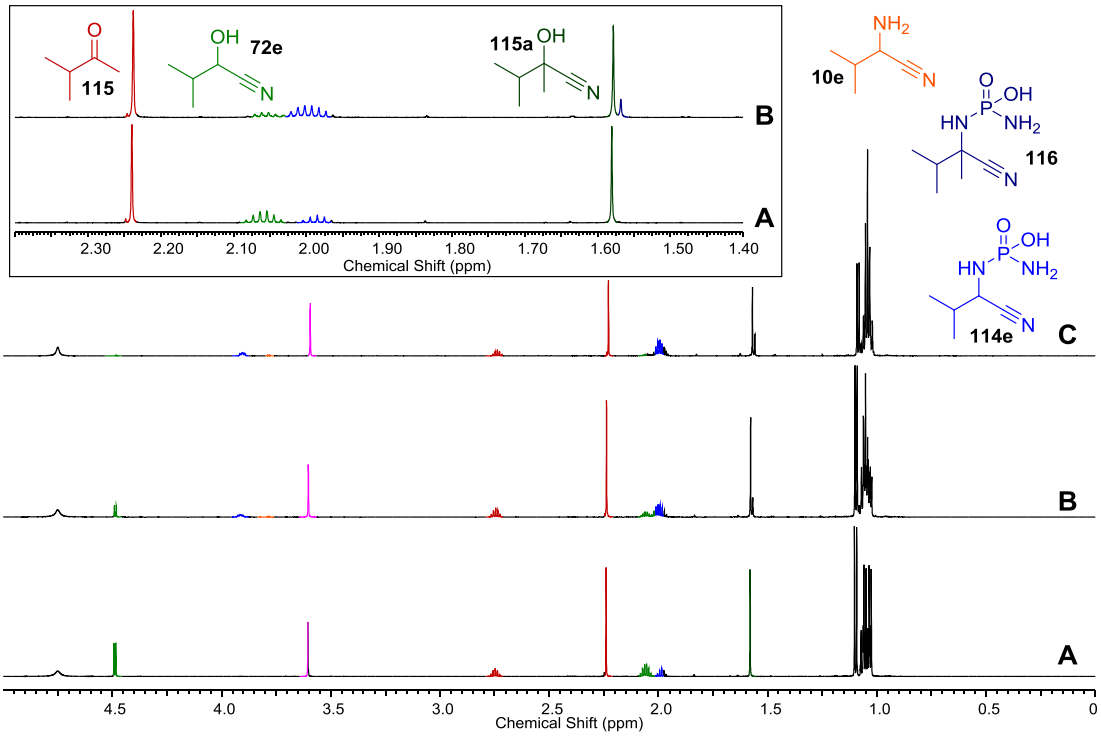


Figure 7.2.7: *Water-supressed ^{1}H NMR spectra (700 MHz, H_2O:D_2O 9:1, 0.00–5.00 ppm; inset 1.40-2.40) showing the reaction of isobutyraldehyde (200 mM) and 3-methyl-butanone (200 mM) with sodium cyanide (1.2 eq.) and DAP (4 eq) at pH 7 at room temperature after:* ***A****) 30 minutes,* ***B****) 1 d, and* ***C****) 4 d.*

7.2.5 Exploring the reaction scope

These results are summarised in ***Table 2.2***.

2-Methylbutyraldehyde ***8f*** *(R=CH(CH_3)CH_2CH_3)*

t(d)	8f	72f	114f	10f	114f (by ^{31}P)	% aldehyde observed
0	7%	-	16%	5%	10%	28%
1	8%	-	34%	9%	27%	51%
2	8%	-	45%	8%	40%	61%
3	8%	-	51%	9%	55%	68%
4	10%	-	62%	8%	66%	80%
6	8%	-	72%	8%	76%	88%
8	8%	-	83%	9%	76%	100%

Table 7.2.34: *reaction of 2-methylbutyraldehyde* ***8f*** *(200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature at pH 7, general procedure 2.1.*

Diagnostic NMR data for species observed in the phosphoro-Strecker reaction of 2-methylbutyraldehyde **8f**:

8f: ^{1}H NMR (700 MHz,D_2O) δ_H 0.85 (3H, t, J = 7.4 Hz).

10f (two diastereomers): ^{1}H NMR (700 MHz,D_2O) δ_H 3.83 (d, J = 5.3 Hz, (C2)-H), 3.89 (d, J = 5.5 Hz, (C2)-H).

114f (two diastereomers): ^{1}H NMR (700 MHz,D_2O) δ_H 4.0 (1H, dd, *J* = 7.2, 5.9 Hz, (C2)-H), 3.95 (1H, dd, *J* = 7.2, 5.9 Hz, (C2)-H), 1.79 – 1.71 (2H, m, (C3)-H), 1.59 (1H, qdd, *J* = 13.5, 7.5, 4.7 Hz, (C4)-H), 1.52 (1H, qdd, *J* = 13.7, 7.2, 5.5 Hz, (C4)-H′), 1.24 (2H, m, (C4)-H), 1.0 (6H, dd, J = 6.9, 3.3 Hz, (C3)-CH_3), 0.91 (6H, td, *J* = 7.5, 1.2 Hz, (C5)-H_3). ^{31}P NMR (264 MHz,D_2O) δ_P 10.95 (d, *J* = 9.5 Hz), 10.87 (d, *J* = 8.6 Hz).

3-Methylbutyraldehyde ***8g*** *(R=$CH_2CH(CH_3)_2$)*

t (d)	8g	72g	114g	126g	10g	114g (by ^{31}P)	126g (by ^{31}P)	% aldehyde observed
0	4%	-	10%	13%	0%	9%	7%	27%
1	4%	-	56%	6%	2%	62%	2%	68%
2	3%	-	61%	6%	2%	70%	2%	72%
3	4%	-	69%	6%	2%	74%	3%	81%
4	4%	-	70%	6%	2%	71%	8%	82%

Table 7.2.35: *reaction of 3-methylbutyraldehyde* ***8g*** *(200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature at pH 7, general procedure 2.1.*

Diagnostic NMR data for species observed in the phosphoro-Strecker reaction of 3-methylbutyraldehyde **8g**:

8g: ^{1}H NMR (600 MHz,D_2O) δ_H 9.58 (1H, t, *J* = 2.5 Hz, (C1)-H), 2.31 (2H, dd, *J* = 6.9, 2.5 Hz, (C2)-H), 2.09 (1H, sept, *J* = 6.8 Hz, (C3)-H), 0.86 (6H, app d, *J* = 6.7 Hz, (C4)-H_3 and (C3)-CH_3 overlapping).

10g: ^{1}H NMR (700 MHz,D_2O) δ_H 3.87 (1H, dd, *J* = 8.7, 6.8 Hz, (C2)-H).

114g: ^{1}H NMR (700 MHz,D_2O) δ_H 4.07 (1H, dt, *J* = 16.0, 8.0 Hz, Y of ABXY), 1.85-1.79 (1H, m, X of ABXY), 1.74-1.69 (1H, m, ABXY), 1.68-1.64 (1H, m, ABXY), 0.93 (6H, d, *J* = 6.7 Hz). ^{31}P NMR (264 MHz,D_2O) δ_P 10.47.

126g: ^{1}H NMR (700 MHz,D_2O) δ_H 4.05-4.00 (1H, m, (C2)-H. ^{31}P NMR (264 MHz,D_2O) δ_P 7.15.

Butyraldehyde ***8h*** *(R=$CH_2CH_2CH_3$)*

t(d)	8h	72h	114h	10h	114h (by ^{31}P)	% aldehyde observed
0	2%	-	5%	0%	5%	9%
1	3%	-	29%	2%	28%	34%
2	4%	-	36%	2%	34%	42%
3	4%	-	39%	2%	36%	45%
4	4%	-	39%	2%	37%	45%

Table 7.2.36: *reaction of butyraldehyde* ***8h*** *(200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature at pH 7, general procedure 2.1.*

Diagnostic NMR data for species observed in the phosphoro-Strecker reaction of butyraldehyde **8h**:

8h: ^{1}H NMR (700 MHz,D_2O) δ_H 0.88 (3H, t, *J* = 7.1 Hz, (C4)-H_3).

10h: ^{1}H NMR (700 MHz,D_2O) δ_H 3.85 (1H, t, *J* = 7.5 Hz, (C2)-H).

114h: ^{1}H NMR (700 MHz,D_2O) δ_H 4.04 (1H, dt, *J* = 15.6, 7.4 Hz, (C2)-H), 1.78-1.74 (2H, m, (C3)-H_2), 1.50-1.44 (2H, m, (C4)-H_2), 0.93 (3H, t, J = 7.8 Hz, (C5)-H_3). ^{31}P NMR (264 MHz,D_2O) δ_P 10.61.

*Pentanaldehyde **8i** (R=$(CH_2)_3CH_3$)*

t(d)	8i	72i	114i	126i	10i	114i (by ^{31}P)	126i (by ^{31}P)	% aldehyde observed
0	5%	-	10%	6%	0%	8%	2%	21%
1	5%	-	45%	5%	3%	44%	1%	58%
2	5%	-	50%	4%	2%	49%	1%	61%
3	5%	-	50%	4%	2%	51%	1%	61%
4	5%	-	50%	4%	1%	53%	1%	61%

Table 7.2.37: *reaction of pentanaldehyde **8i** (200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature at pH 7, general procedure 2.1.*

Diagnostic NMR data for species observed in the phosphoro-Strecker reaction of pentanaldehyde **8i**:

8i: ^{1}H NMR (700 MHz,D_2O) δ_H 2.17 (2H, t, *J* = 7.1 Hz, (C2)-H).

10i: ^{1}H NMR (700 MHz,D_2O) δ_H 3.82 (1H, t, *J* = 6.7 Hz, (C2)-H).

114i: ^{1}H NMR (700 MHz,D_2O) δ_H 4.03 (1H, dt, *J* = 16.1, 7.2 Hz, (C2)-H), 1.79 (2H, dt, *J* = 15.2, 7.1 Hz, (C3)-H_2), 1.46-1.42 (2H, m, (C4)-H_2), 1.37-1.33 (2H, m, (C5)-H_2), 0.89 (3H, t, *J* = 7.4 Hz, (C6)-H_3). ^{31}P NMR (264 MHz,D_2O) δ_P 10.61.

126i: ^{1}H NMR (700 MHz,D_2O) δ_H 4.00-3.97 (1H, m, (C2)-H. ^{31}P NMR (264 MHz,D_2O) δ_P 7.36.

*2-Ethylbutyraldehyde **8j** (R=$CH(CH_2CH_3)_2$)*

t(d)	8j	72j	114j	10j	114j (by ^{31}P)	% aldehyde observed
0	6%	-	4%	1%	2%	11%
1	5%	-	14%	2%	7%	21%
2	7%	-	23%	2%	14%	32%
3	6%	-	28%	2%	21%	36%
4	6%	-	34%	2%	26%	42%
5	6%	-	38%	2%	30%	46%
6	6%	-	39%	2%	32%	47%
7	6%	-	43%	2%	33%	51%
8	6%	-	42%	2%	35%	50%

Table 7.2.38: *reaction of 2-ethylbutyraldehyde **8j** (200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature at pH 7, general procedure 2.1.*

Diagnostic NMR data for species observed in the phosphoro-Strecker reaction of 2-ethylbutyraldehyde **8j**:

8j: ^{1}H NMR (700 MHz,D_2O) δ_H 0.84 (3H, t, *J* = 7.5 Hz, (C4)-H_3).

10j: ^{1}H NMR (700 MHz,D_2O) δ_H 3.94 (1H, t, *J* = 5.0 Hz, (C2)-H).

114j: ^{1}H NMR (700 MHz,D_2O) δ_H 4.05 (1H, dd, *J* = 9.0, 5.8 Hz, (C2)-H), 1.56-1.49 (m), 1.48-1.41 (m), 1.39-1.33 (m), 0.89-0.86 (6H, m). ^{31}P NMR (264 MHz,D_2O) δ_P 10.99.

*Octanal **8k** ($R=(CH_2)_6CH_3$)*

MeCN (100 µL), 100 mM pentaerythritol standard (50 µL), 0.42 M NaCN stock solution (150 µL, 0.06 mmol) and octanal **8k** (8.0 µL, 0.05 mmol) were placed into 10 Eppendorf tubes equipped with stirrer bars. 0.82 M DAP stock solution (250 µL, 0.20 mmol) adjusted to pH 7 was then added to each Eppendorf. After the desired reaction time, the pH of the reaction was monitored, the solution lyophilised and resuspended in D_2O for NMR analysis.

t(d)	**114k**	**114k** (by ^{31}P)	pH
0	6%	6%	7.5
1	15%	13%	8.0
2	17%	16%	8.0
4	31%	28%	8.1
5	35%	31%	8.3
7	34%	33%	8.3
8	32%	32%	8.4
10	35%	33%	8.6

Table 7.2.39: *reaction of octanal **8k** (100 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in 4:1 H_2O:MeCN at room temperature at pH 7.*

114h: ^{1}H NMR (700 MHz,D_2O) δ_H 4.03 (1H, dt, *J* = 15.8, 8.4 Hz, (C2)-H), 1.78 (2H, dt, *J* = 16.3, 7.8 Hz, (C3)-H_2), 1.49-1.44 (2H, m, (C4)-H_2), 1.36-1.26 (8H, m, (C5)-H_2, (C6)-H_2, (C7)-H_2 and (C8)-H_2 overlapping), 0.86 (3H, t, *J* = 6.9 Hz, (C9)-H_3). ^{31}P NMR (264 MHz,D_2O) δ_P 10.61.

*Lactonitrile **72b***

Procedure (omitting NaCN): lactonitrile **72b** (36.9 mg, 0.52 mmol) was added to 100 mM pentaerythritol standard (250 µL) and water (1 mL). DAP (MW = 173.48, 344.58 mg, 1.99 mmol) was added, the solution adjusted to pH 7.1, and the total volume made up to 2.5 mL. The pH was monitored over the course of the reaction and periodically adjusted back to pH 7. 500 µL aliquots were periodically removed, diluted with 100 µL D_2O, and NMR spectra collected. Yields shown in ***Table 7.2.40.***

Diagnostic NMR data for species observed:

114b: ^{1}H NMR (400 MHz,D_2O) δ_H 4.17-4.10 (1H, q, *J* = 6.8 Hz, (C2)-H), 1.48 (3H, d, *J* = 7.0 Hz, (C3)-H). ^{31}P NMR (162 MHz,D_2O) δ_P 10.50.

72b: ^{1}H NMR (400 MHz,D_2O) δ_H 1.52 (3H, d, *J* = 7.0 Hz, (C3)-H).

10b: ^{1}H NMR (400 MHz,D_2O) δ_H δ 1.34 (3H, d, *J* = 5.3 Hz, (C3)-H).

t(d)	72b	114b	10b	114b (by ^{31}P)	% aldehyde observed by integration of all methyl peaks
0	72%	9%	1%	7%	83%
1	2%	84%	2%	68%	91%
2	2%	90%	1%	68%	95%
3	2%	89%	1%	72%	92%
4	2%	84%	2%	71%	94%

Table 7.2.40: *reaction of lactonitrile* ***72b*** *(200 mM) with DAP (4.0 eq) in H_2O at room temperature at pH 7.*

Procedure (using NaCN): lactonitrile **72b** (36.9 mg, 0.52 mmol) was added to 100 mM pentaerythritol standard (250 µL) and water (1 mL). NaCN (5.4 mg, 0.11 mmol) and DAP (MW = 173.48, 344.58 mg, 1.99 mmol) were added, the solution adjusted to pH 7.0, and the total volume made up to 2.5 mL. The pH was monitored over the course of the reaction and periodically adjusted back to pH 7. 500 µL aliquots were periodically removed, diluted with 100 µL D_2O, and NMR spectra collected. *Yields shown in* ***Table 7.2.41.***

t(d)	72b	114b	10b	114b (by ^{31}P)	% aldehyde observed by integration of all methyl peaks
0	76%	9%	1%	7%	88%
1	4%	85%	2%	69%	96%
2	5%	86%	1%	72%	92%
3	1%	89%	1%	74%	94%
4	1%	88%	1%	74%	94%

Table 7.2.41: *reaction of lactonitrile* ***72b*** *(200 mM) with NaCN (0.2 eq) and DAP (4.0 eq) in H_2O at room temperature at pH 7.*

Acetaldehyde ***71***

t(d)	71	72b	114b	10b	114b (by ^{31}P)	% aldehyde observed by integration of all methyl peaks
0	-	38	3	0	3%	49
1	-	6	35	2	34%	49
2	-	2	40	2	39%	50
3	-	2	49	2	40%	65
4	-	2	48	1	42%	58
5	-	0	51	1	41%	60

Table 7.2.42: *reaction of acetaldehyde* ***71*** *(100 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature at pH 7, general procedure 2.2.*

t(d)	71	72b	114b	10b	114b (by ^{31}P)	% aldehyde observed by integration of all methyl peaks
0	-	49	5	0	7	61
1	-	11	47	2	32	71
2	-	3	47	2	38	61
3	-	2	48	2	39	58
4	-	2	56	1	41	69

Table 7.2.43: *reaction of acetaldehyde* ***71*** *(200 mM) with NaCN (1.2 eq) and DAP (2.0 eq) in H_2O at room temperature at pH 7, general procedure 2.2.*

Acrolein ***121***

Attempt 1: Acrolein **121** (11.1 mg, 0.198 mmol) and NaCN (34.9 mg, 0.71 mmol) were dissolved in water (400 µL) and 100 mM pentaerythritol standard (100 µL) added. The mixture was adjusted to pH 9.1. After 30 minutes, DAP (MW = 195.81, 156.06 mg, 0.797 mmol) in water (400 µL) was added and the solution adjusted to pH 6.9. The solution was diluted with D_2O (100 µL), transferred to an NMR tube and periodically subjected to NMR analysis, which revealed the formation of a complex mixture of products; the desired *N*-phosphordiamidate amino nitrile could not be detected.

Attempt 2: **121** (13.1 mg, 0.234 mmol) and NaCN (39.1 mg, 0.798 mmol) were dissolved in water (400 µL) and 100 mM pentaerythritol standard (100 µL) added. The mixture was adjusted to pH 9.2 and made up to 800 µL with water. After 30 minutes, an aliquot (400 µL) was withdrawn, diluted with D_2O (100 µL) and subjected to NMR analysis, which revealed the formation of a complex mixture of products. DAP (MW = 195.81, 91.48 mg, 0.467 mmol) in water (400 µL) was added and the solution adjusted to pH 6.9. The solution was diluted with D_2O (100 µL), transferred to an NMR tube and periodically subjected to NMR analysis, which revealed the formation of a complex mixture of products; the desired *N*-phosphordiamidate amino nitrile could not be detected.

*Glycolaldehyde **32***

Diagnostic NMR peak for **114c:** ^{31}P NMR (162 MHz,D_2O) δ_P 10.67.

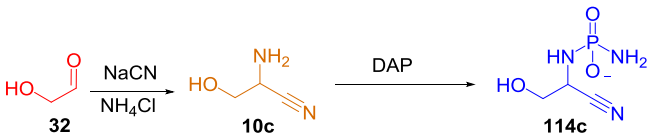
		% yield of **114c** by relative integration ^{31}P NMR					
	DAP stoichiometry	1	2	4	1	2	4
	initial pH	9.6	9.5	9.5	9.1	9.1	8.9
time/d	0	0%	0%	0%	1%	3%	4%
	1	1%	1%	3%	4%	9%	13%
	2	-	-	-	7%	16%	23%
	3	-	-	-	10%	21%	29%
	6	4%	9%	14%	17%	24%	37%

***Table 7.2.44**: reaction of glycolaldehyde **32** (0.1 mmol, 200 mM) with NaCN (1.2 eq) and DAP (1-4 eq) in D_2O.*

		% yield of **114c** by relative integration ^{31}P NMR				
	initial pH	11.0	9.9	8.9	7.9	7.0
Time / d	0	0%	0%	0%	0%	0%
	1	2%	2%	2%	3%	3%
	2	4%	4%	6%	5%	6%
	5	5%	7%	7%	8%	9%
	7	5%	7%	8%	10%	10%
	13	6%	9%	11%	10%	11%
	19	6%	10%	11%	11%	12%
	final pH	10	9.9	9.5	9.5	9.0

***Table 7.2.45**: reaction of glycolaldehyde **32** (50 mM) with NaCN (1.2 eq) and DAP (4 eq) in D_2O.*

*Formation of glycolaldehyde aminonitrile **10c** and its reaction with DAP*



Glycolaldehyde stock solution (3.87 M, 310 µL) was added to NaCN stock solution (0.8 M, 1.74 mL), followed by NH_4Cl stock solution (4.6 M, 640 µL). The reaction mixture was adjusted to pH 9.5 and stirred at room temperature. After 5 days, aliquots of the reaction (270 µL) were taken and added to DAP stock solution (1.83 M), diluted with D_2O to 500 µL and adjusted to pH 10. Yields shown in ***Table 7.2.46.***

	% yield of **114c** by relative integration ^{31}P NMR			
time/d	1 eq DAP	1.5 eq DAP	2 eq DAP	4 eq DAP
0	0%	0%	0%	0%
2	0%	0%	<1%	1%

Table 7.2.46: *reaction of glycolaldehyde aminonitrile (200 mM) and DAP in D_2O, initial pH 10.0. All yields by relative integration of ^{31}P NMR spectra; no change seen by ^{1}H NMR.*

*2-Methylbutyraldehyde **8f**, forming imine first*

Aldehyde (1 eq) was dissolved in D_2O and pentaerythritol standard and DAP solution (4 eq) were added. NaCN (1.2 eq) was then added, and the solution adjusted to pH 7.0. The pH was monitored over the course of the reaction (and periodically adjusted back to pH 7). 500 µL aliquots were periodically removed, freeze-dried, resuspended in D_2O and NMR spectra collected. Yields for the reverse addition of substrates (forming cyanohydrin first) are shown in ***Table 7.2.34***.

t(d)	**8f**	**72f**	**114f** + **10f***	**114f** (by ^{31}P)	% aldehyde observed	
					sum of known peaks	by integration of methyl peaks
0	5%	-	0%	0%	5%	10%
1	3%	-	1%	4%	4%	5%
2	5%	-	1%	3%	6%	9%
3	5%	-	1%	4%	6%	8%
4	4%	-	1%	3%	5%	8%
6	4%	-	1%	4%	5%	9%
8	3%	-	1%	5%	4%	11%

Table 7.2.47: *reaction of 2-methylbutyraldehyde **8f** (200 mM) with NaCN (1.2 eq) and DAP (4.0 eq) in H_2O at room temperature with the reaction mixture periodically adjusted back to pH 7.0.*
** peaks overlap*

7.2.6 Preparative synthesis and purification of *N*-phosphordiamidate aminonitrile 114

General procedure 2.3: Aldehyde (1 mmol) was stirred in water (1.5 mL), then NaCN (1.3 mmol) was added, and the solution adjusted to pH 9. DAP (4.0 mmol) was dissolved in water (1.5 mL) and added to the cyanohydrin mixture. The resulting mixture was adjusted to pH 7.1, made up to 5 mL with water, and stirred at room temperature, with daily adjustment to pH 7. The reaction was followed by ^{31}P NMR analysis. Purification by several rounds of precipitation took place, with the supernatant concentrated *in vacuo* in a water bath set to 25 °C after each step. The precipitate was lyophilised, dissolved in D_2O and subjected to NMR analysis; an aliquot of the supernatant was also diluted with D_2O and subjected to NMR analysis. Relative

integration of all species detectable by ^{31}P NMR meant that the presence of DAP and product in the precipitate and supernatant could be monitored.

Isoleucine N-phosphordiamidate aminonitrile ***114f***

2-methylbutyraldehyde **8f** (82.8 mg, 0.96 mmol), NaCN (63.3 mg, 1.29 mmol), DAP (782.6 mg, 4.0 mmol), general procedure 2.3. The reaction appeared complete after 11 d (when **114f** made up 11% of all species detectable by ^{31}P NMR). Precipitation steps shown in ***Table 7.2.48***. The supernatant from the last step was concentrated *in vacuo*, lyophilised, and dissolved in water (2 mL). An aliquot of this solution was spiked with pentaerythritol: a total yield of 0.36 mmol (38%) was calculated from ^{1}H NMR integral analysis. (Relative integration of all species observed by ^{31}P NMR analysis indicated that the product made up 79% of all phosphorylated species).

114f (two diastereomers): ^{1}H NMR (700 MHz,D_2O) δ_H 4.0 (1H, dd, *J* = 7.2, 5.9 Hz, (C2)-H), 3.95 (1H, dd, *J* = 7.2, 5.9 Hz, (C2)-H), 1.79 – 1.71 (2H, m, (C3)-H), 1.59 (1H, qdd, *J* = 13.5, 7.5, 4.7 Hz, (C4)-H), 1.52 (1H, qdd, *J* = 13.7, 7.2, 5.5 Hz, (C4)-H′), 1.24 (2H, m, (C4)-H), 1.0 (6H, dd, J = 6.9, 3.3 Hz, (C3)-CH_3), 0.91 (6H, td, *J* = 7.5, 1.2 Hz, (C5)-H_3). ^{13}C NMR (150 MHz,D_2O) δ_C 122.7 (d, J = 1.7 Hz, C1), 122.2 (d, J = 2.8 Hz, C1), 49.5 (C2), 49.4 (C2), 39.5 (d, *J* = 6.6 Hz, C3), 39.3 (d, *J* = 6.6 Hz, C3), 26.1 (C4), 25.3 (C4), 15.3 ((C3)-CH_3), 14.9 ((C3)-CH_3), 11.32 (C5), 11.31 (C5). IR (cm^{-1}) 2965-2875 (C-H), 2231 (nitrile C≡N), 1563, 1460, 1413, 1180, 1049, 869. HRMS (*m/z*): $[M+H^+]^+$ calcd for formula $C_6H_{14}N_3O_2P^+$, 192.0902; found 192.0903.

Step	Aqueous layer volume	Precipitation solvent, volume	Precipitation temp, time	Precipitate composition		Supernatant composition	
				DAP	**114f**	DAP	**114f**
1	5 mL	EtOH, 8mL	0 °C, 20 mins	52%	0%	-	-
2	4 mL	MeOH, 8 mL	- 10 °C, 20 mins	68%	32%	68%	32%
3	2 mL	EtOH, 10 mL	- 10 °C, 20 mins	87%	<1%	40%	60%
4	2 mL	EtOH, 10 mL	- 10 °C, 20 mins	86%	<1%	0%	93%

Table 7.2.48: *successful purification of* ***114f*** *by selective precipitation of residual DAP. Centrifuged for 5 mins, 3000 rpm at each step.*

Valine N-phosphordiamidate aminonitrile ***114e***

Isobutyraldehyde **8e** (74.0 mg, 1.03 mmol), NaCN (99.6 mg, 2.03 mmol), DAP (783.6 mg, 4.0 mmol), general procedure 2.3. The reaction appeared complete after 11 d (when the desired product made up 12% of all species detectable by ^{31}P NMR). Precipitation steps shown in ***Table 7.2.49.*** The supernatant from the last step was concentrated *in vacuo*, lyophilised, and dissolved in water (2 mL). An aliquot of this solution was spiked with pentaerythritol: a total yield of 0.32 mmol (31%) was

calculated from ^{1}H NMR integral analysis. (Relative integration of all species detectable by ^{31}P NMR indicated that the product made up 79% of all phosphorylated species).
114e: ^{1}H NMR (700 MHz,D_2O) δ_H 3.89 (1H, dd, *J* = 6.1, 9.0 Hz, (C2)-H),1.98 (1H, app oct, *J* = 6.7, (C3)-H), 1.03 (6H, app t, *J* = 6.7 Hz, (C4)-H_3 and (C4′)-H_3 overlapping). ^{13}C NMR (150 MHz,D_2O) δ_C 122.3 (d, *J* = 1.7 Hz, C1), 50.9 (C2), 33.0 (d, *J* = 6.6 Hz, C3), 18.8 (C4), 17.9 (C4′). ^{31}P NMR (264 MHz,D_2O) δ_P 10.88 (d, *J* = 8.6 Hz). IR (cm^{-1}) 3410-3280, 2970-2875, 1577. HRMS (*m/z*): $[M+H^+]^+$ calcd for formula $C_5H_{12}N_3O_2P^+$, 178.0725; found 178.0745.

Step	Aqueous layer volume	Precipitation solvent, volume	Precipitation temp, time	Precipitate composition		Supernatant composition	
				DAP	**114e**	DAP	**114e**
1	5 mL	EtOH, 8 mL	0 °C, 20 mins	55%	0%	-	-
2	4 mL	MeOH, 8 mL	- 10 °C, 20 mins	11%	5%	63%	37%
3	2 mL	EtOH, 10 mL	- 10 °C, 20 mins	86%	4%	26%	74%
4	2 mL	EtOH, 10 mL	- 10 °C, 20 mins	71%	14%	0%	92%

Table 7.2.49: *successful purification of* ***114d*** *by selective precipitation of residual DAP. Centrifuged for 5 mins, 3000 rpm at each step.*

Alanine N-phosphordiamidate aminonitrile ***114b***

Lactonitrile **72b** (71.8 mg, 1.01 mmol) was stirred in water (3.5 mL), and adjusted from pH 12.3 to pH 9.3. DAP (781.3 mg, 4.0 mmol) was added. The resulting mixture was adjusted to pH 6.9, made up to 5 mL with water, and stirred at room temperature, with daily adjustment to pH 7. The reaction was followed by ^{31}P NMR analysis. The reaction was complete after 3 d (when the desired product made up 17% of all species detectable by ^{31}P NMR). Precipitation steps shown in ***Table 7.2.50***. The supernatant from the final step was concentrated *in vacuo*, lyophilised, and dissolved in water (500 µL). An aliquot of this solution was spiked with pentaerythritol: a total yield of 0.598 mmol (59%) was calculated from ^{1}H NMR integral analysis. (Relative integration of all species detectable by ^{31}P NMR indicated that the product made up 72% of all phosphorylated species).
114b: ^{1}H NMR (700 MHz,D_2O) δ_H 4.11 (1H, app quin, *J* = 7.4 (C2)-H), 1.48 (3H, d, *J* = 7.2 Hz, (C3)-H_3). ^{13}C NMR (150 MHz,D_2O) δ_C 123.5 (C1), 39.6 (C2), 21.4 (d, *J* = 7.2 Hz, C3). ^{31}P NMR (700 MHz,D_2O) δ_P 10.30 (d, *J* = 7.6 Hz). IR (cm^{-1}) 3410-3280, 2349**.** HRMS (*m/z*): $[M+Na^+]^+$ calcd for formula $C_3H_7N_3O_2PNa^+$, 172.0252; found 172.0244.

Step	Aqueous layer volume	Precipitation solvent, volume	Precipitation temperature, time	precipitate		supernatant	
				DAP	**114b**	DAP	**114b**
1	5 mL	MeOH, 6 mL	-10 °C, 20 mins	39%	12%	-	-
2	11 mL[a]	EtOH, 5 mL	- 10 °C, 20 mins	63%	1%	79%	20%
3	4 mL	EtOH, 4 mL + MeOH, 4 mL	- 10 °C, 20 mins	89%	3%	72%	28%
4	2 mL	EtOH, 5.5 mL+ MeOH, 5.5 mL	- 10 °C, 20 mins	93%	2%	55%	42%
5	1.5 mL	EtOH, 6 mL + MeOH 4 mL	- 10 °C, 20 mins	80%	2%	42%	55%
6	1 mL	EtOH, 8 mL + MeOH 3 mL	- 10 °C, 20 mins	80%	3%	26%	71%
7	1.5 mL	EtOH, 10 mL + MeOH, 1.5 mL	- 10 °C, 20 mins	70%	9%	17%	73%
8	1.5 mL	EtOH, 10 mL + MeOH, 1.5 mL	- 10 °C, 20 mins	20%	15%	20%	80%
9	1.5 mL	EtOH, 10 mL + MeOH, 1.5 mL	- 10 °C, 20 mins	42%	58%	17%	73%

Table 7.2.50: *successful purification of* ***114b*** *by selective precipitation of residual DAP. Centrifuged for 5 mins, 2500 rpm at each step.*

[a] *the supernatant from step 1 was used without concentration*

7.2.7 Derivatisation experiments

7.2.5.1 Acid hydrolysis of *N*-phosphordiamidate aminonitrile 114

General procedure 2.4: A stock solution of either purified or unpurified *N*-phosphoroaminonitrile **114** (0.05 mmol) was adjusted to pH 1.5 with the addition of 4 M HCl, diluted with water to a total volume of 500 µL and left at room temperature or heated at 50 °C. The reaction mixture was periodically readjusted to pH 1.5 and aliquots (100 µL) were removed, diluted with D_2O(400 µL) and subjected to NMR analysis.

Acid hydrolysis of alanine N-phosphordiamidite aminonitrile ***114b***

Characterisation data for the products observed:

10b: ^{1}H NMR (600 MHz, 9:1 H_2O:D_2O) δ_H ^{1}H (600 MHz, 9:1 H_2O:D_2O) δ 4.52 (1H, q, *J* = 7.1 Hz, (C2)-H), 1.64 (3H, d, *J* = 7.0 Hz, (C3)-H_3). ^{13}C NMR (150 MHz, 9:1 H_2O:D_2O) δ_C 117.5 (C1), 38.0 (C2), 17.2 (C3).

72b: ^{1}H NMR (600 MHz, 9:1 H_2O:D_2O) δ_H ^{1}H (600 MHz, 9:1 H_2O:D_2O) δ 4.65 (1H, q, *J* = 6.9 Hz, (C2)-H), 1.47 (3H, d, *J* = 7.0 Hz, (C3)-H_3). ^{13}C NMR (150 MHz, 9:1 H_2O:D_2O) δ_C 122.3 (C1), 57.3 (C2), 21.0 (C3).

t(h)	**114b**	**10b**
0	100%	0%
4	0%	100%

Table 7.2.51: *acid hydrolysis of purified alanine N-phosphoroaminonitrile* ***114b*** *at 50 °C. Yields shown are calculated by relative integration of identifiable species seen by* ^{1}*H NMR.*

t(h)	**114b** (as % of all phosphorylated species by ^{31}P)	**114b**	**10b**	**72b**
0	20%	100%	0%	0%
4	6%	-	-	-
8	0%	0%	77%	23%

Table 7.2.52: *acid hydrolysis of unpurified alanine N-phosphoroaminonitrile* ***114b*** *at 50 °C. Yields shown are calculated by relative integration of identifiable species seen by* ^{1}*H NMR.*

t(h)	**114b**	**10b**
0	96%	4%
4	79%	21%
8	65%	35%
12	39%	61%
16	29%	71%
22	21%	79%
30	10%	90%

Table 7.2.53: *acid hydrolysis of unpurified alanine N-phosphoroaminonitrile* ***114b*** *at room temperature. Yields shown are calculated by relative integration of identifiable species seen by* ^{1}*H NMR.*

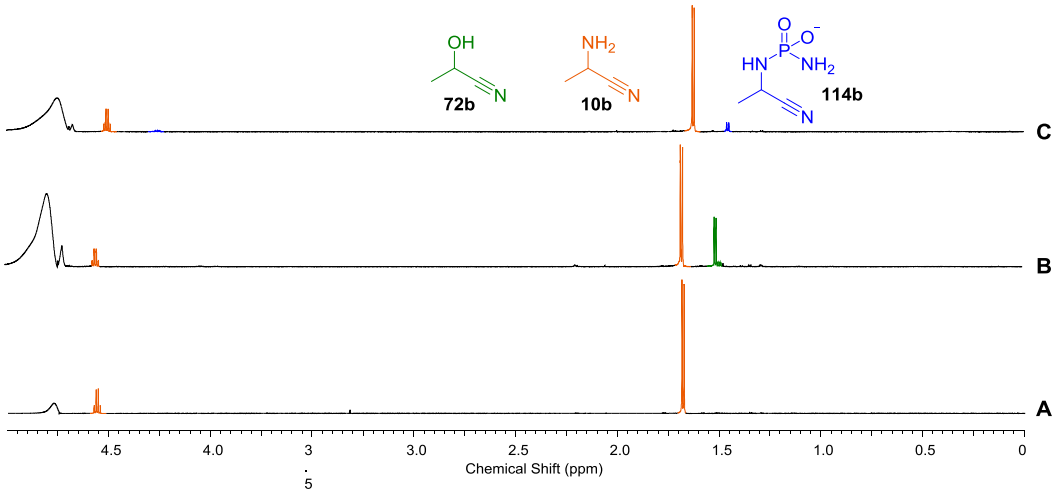


Figure 7.2.7: *Water supressed* ^{1}*H NMR spectra (*H_2O*:*D_2O *1:9, 0.00-5.00 ppm, 700 MHz) showing the hydrolysis of alanine N-phosphoroaminonitrile* ***114b*** *(100 mM) at pH 1.5:* ***A****) after 4 h at 50 °C, purified* ***114b*** *cleanly converts to* ***10b****;* ***B****) after 8 h at 50 °C, unpurified* ***114b*** *converts to* ***10b*** *(77% by relative integration identifiable species) and* ***72b*** *(23%);* ***C****) after 30 h at room temperature, unpurified* ***114b*** *has undergone near-complete clean conversion to* ***10b*** *(90% by relative integration identifiable species). The slight difference in chemical shift of the aminonitrile methyl protons in spectrum C is due to a difference in pH between the samples.*

*Acid hydrolysis of isoleucine N-phosphordiamidite aminonitrile **114f***

Characterisation data for the products observed:

10f, diastereomer A: ^{1}H NMR (600 MHz, 9:1 H_2O:D_2O) δ_H 4.47 (1H, d, *J* = 5.1 Hz, (C2)-H), 2.05-1.98 (1H, m, overlaps with diastereomer B, (C3)-H), 1.51-1.52 (1H, m, (C4)-H), 1.38-1.31 (1H, m, (C4)-H′), 1.06 (3H, d, *J* = 6.7 Hz, (C3)-CH_3), 0.88 (1H, t, *J* = 7.4 Hz, (C5)-H_3). ^{13}C NMR (150 MHz, 9:1 H_2O:D_2O) δ_C115.3 (C1), 47.0 (C2), 36.5 (C3), 26.3 (C4), 14.3 (C3)-CH_3), 10.9 (C5).

10f, diastereomer B: ^{1}H NMR (600 MHz, 9:1 H_2O:D_2O) δ_H 4.39 (1H, d, *J* = 6.1 Hz), 2.05-1.98 (1H, m, overlaps with diastereomer A), 1.59 (1H, ddd, *J* = 13.5, 7.4, 4.0 Hz, (C4)-H), 1.31-1.23 (1H, m, (C4)-H′),), 1.04 (3H, d, *J* = 7.0 Hz, (C3)-CH_3), 0.90 (1H, t, *J* = 7.2 Hz, (C5)-H_3). ^{13}C NMR (150 MHz, 9:1 H_2O:D_2O) δ_C116.1 (C1), 47.3 (C2), 36.5 (C3), 24.6 (C4), 15.0(C3)-CH_3), 11.0 (C5).

72f: ^{1}H NMR (600 MHz, 9:1 H_2O:D_2O) δ_H 4.52 (1H, app d, partially obscured by **10f**, diastereomer A, (C2)-H), 1.82-1.73 (1H, (C3)-H). ^{13}C NMR (150 MHz, 9:1 H_2O:D_2O) δ_C122.7 (C1), 39.5 (C3).

t(h)	**114f** (as % of all phosphorylated species by ^{31}P)	**114f**	**10f**
0	78%	85%	5%
4	26%	61%	39%
22	0%	11%	89%
24	0%	10%	90%

***Table 7.2.54:** acid hydrolysis of purified isoleucine N-phosphoroaminonitrile **114f** at 50 °C. Yields shown are calculated by relative integration of identifiable species seen by ^{1}H NMR.*

t(h)	**114f** (as % of all phosphorylated species)	**114f** + **72f***	**10f**
0	14%	95%	5%
4	7%	80%	20%
8	<1%	22%	78%
12	0%	14%	86%

***Table 7.2.55:** acid hydrolysis of unpurified isoleucine N-phosphoroaminonitrile **114f** at 50 °C. Yields shown are calculated by relative integration of identifiable species seen by ^{1}H NMR. A second reaction prepared on the same scale was spiked with pentaerythritol and subjected to NMR analysis after 12 h at 50 °C: this indicated yields of 50% (0.025 mmol) **10f** and 9% (0.005 mmol) **72f**; 64% of initial **114f** could be accounted for by integration of all peaks in methyl region. * **114f** and **72f** peaks overlap.*

*Acid hydrolysis of valine N-phosphordiamidite aminonitrile **114e***

Characterisation data for the products observed:

10e: ^{1}H NMR (600 MHz, 9:1 H_2O:D_2O) δ_H 4.37 (1H, d, *J* = 5.8 Hz, (C2)-H), 2.24 (1H, dq, *J* = 13.1, 6.6 Hz, (C3)-H), 1.07 (3H, d, *J* = 6.7 Hz, (C4)-H_3), 1.05 (3H, d, *J* = 6.7 Hz, (C4′)-H_3). ^{13}C NMR (150 MHz, 9:1 H_2O:D_2O) δ_C 115.8 (C1), 48.2 (C2), 30.4 (C3), 18.7 (C4), 17.2 (C4′).

72e: ^{1}H NMR (600 MHz, 9:1 H_2O:D_2O) δ_H 4.38 (1H, app d, partially obscured by **10e**, (C2)-H), 1.95 (1H, dq, *J* = 12.9, 6.6 Hz, (C3)-H), 0.95 (1H, d, J = 6.7 Hz, (C4)-H_3), 0.92 (1H, d, *J* = 6.7 Hz, (C4′)-H_3). ^{13}C NMR (150 MHz, 9:1 H_2O:D_2O) δ_C 121.0 (C1), 67.2 (C2), 33.1 (C3), 17.7 (C4′), 17.3 (C4).

t(h)	**114e** (as % of all phosphorylated species by ^{31}P)	**114e**	**10e**	**72e**
0	77%	95%	5%	-
4	15%	-*	46%	-*
22	0%	-	85%	15%
24	0%	-	84%	16%

Table 7.2.55: *acid hydrolysis of purified valine N-phosphoroaminonitrile **114e** at 50 °C. Yields shown are calculated by relative integration of identifiable species seen by ^{1}H NMR.*

** **114e** and **72e** peaks overlap with each other and other unidentified species at 4 h (**114e**+**72e**+others = 54%). Integrating one discrete half of **72e** suggests a yield of 8% **72e**.*

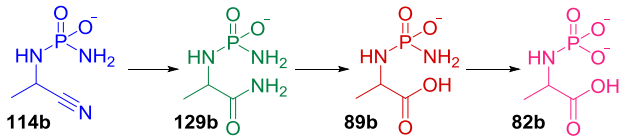
t(h)	**114e*** + others	**10e**	**72e**
0	93%	7%	0%
4	78%	13%	9%
8	49%	39%	12%
12	22%	62%	15%
16	0%	75%	25%

Table 7.2.56: *acid hydrolysis of unpurified **114e** at 50 °C. Yields shown are calculated by relative integration of identifiable species seen by ^{1}H NMR. A second reaction prepared on the same scale was spiked with pentaerythritol and subjected to NMR analysis after 16 h at 50 °C: this indicated yields of 52% **10e** and 20% **72e;** 75% of initial **114e** could be accounted for by integration of all peaks in methyl region. * the consumption of **114e** cannot be followed by analysis of ^{31}P NMR spectra as it overlaps with the DAP peak.*

7.2.5.2 Alkaline hydrolysis of *N*-phosphordiamidate aminonitrile 114

General procedure 2.5: A stock solution of either purified or unpurified *N*-phosphoroaminonitrile **114** (0.05 mmol) was adjusted to pH 13.0 with the addition of 4 M NaOH, diluted with water to a total volume of 500 µL and left at room temperature or heated at 50 °C. Aliquots (100 µL) were periodically removed, diluted with water (400 µL) and subjected to NMR analysis.

*Hydrolysis of alanine N-phosphordiamidate aminonitrile **114b** in alkaline solution*



Characterisation data for the products observed (all 9:1 H_2O:D_2O):

129b: ^{1}H NMR (700 MHz) δ_H 3.59-3.54 (1H, m) 1.24 (d, *J* = 7.2 Hz). ^{31}P NMR (284 MHz, 9:1 H_2O:D_2O) δ_P 11.96 (*J* = 10.5 Hz).

89b: ^{1}H NMR (700 MHz) δ_H 3.36-3.32 (1H, m), 1.02 (d, *J* = 7.0 Hz). ^{13}C NMR (150 MHz) δ_C 184.0. ^{31}P NMR (284 MHz) δ_P12.62 (*J* = 9.5 Hz).

82b: ^{1}H NMR (700 MHz) δ_H 3.37 (1H, dq, *J* = 9.9, 7.0 Hz), 1.10 (3H, d, *J* = 7.0 Hz). ^{13}C NMR (150 MHz) δ_C 184.8 (d, *J* = 7.2 Hz), 53.5 (d, *J* = 8.8 Hz), 22.2 (dd, *J* = 8.3, 4.4 Hz). ^{31}P NMR (284 MHz) δ_P 8.80 (d, *J* = 9.5 Hz).

	114b	129b	89b	82b	DAP	MAP
At pH 6.5	72%	0%	0%	0%	17%	0%
4 h	57%	13%	3%	0%	22%	5%
12 h	21%	15%	26%	8%	25%	6%
1 d	0%	0%	59%	7%	29%	5%
4 d	0%	0%	50%	15%	29%	7%
8 d	0%	0%	40%	26%	28%	6%
13 d	0%	0%	28%	34%	28%	10%
18 d	0%	0%	23%	38%	28%	11%
23 d	0%	0%	18%	49%	28%	5%
33 d	0%	0%	10%	61%	28%	2%

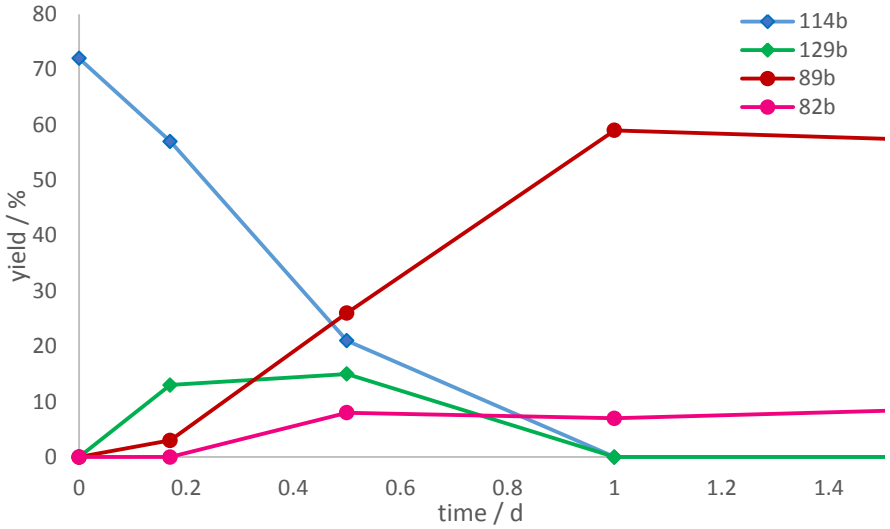
Table 7.2.57:** basic hydrolysis of purified alanine N-phosphoroaminonitrile **114b** at room temperature. Yields shown are calculated by relative integration of identifiable species seen by ^{31}P NMR. These results are summarised in **Graph 2.5**, and in **Graph 7.2.4.

	114b	89b	82b	DAP	MAP
at pH 6.5	72%	0%	0%	17%	0%
1 d	0%	12%	49%	31%	8%
2 d	0%	5%	58%	31%	7%
4 d	0%	0%	58%	32%	10%

***Table 7.2.58:** basic hydrolysis of purified **114b** at 50 °C. Yields shown are calculated by relative integration of identifiable species seen by ^{31}P NMR.*

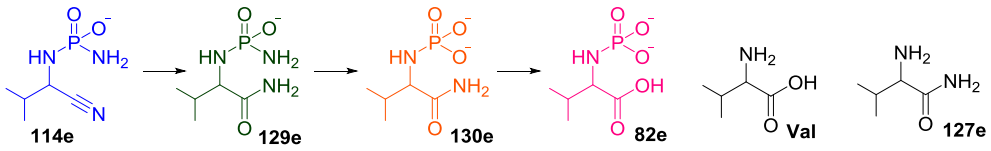
	114b	89b	82b	DAP	MAP
at pH 6.5	20%	0%	0%	76%	0%
1 d	0%	4%	12%	76%	5%
2 d	0%	1%	15%	75%	5%
3 d	0%	<1%	16%	74%	6%

Table 7.2.59: *basic hydrolysis of unpurified* ***114b*** *at 50 °C. Yields shown are calculated by relative integration of identifiable species seen by* ^{31}*P NMR.*



Graph 7.2.4: *Basic hydrolysis of purified* ***114b*** *(100 mM) at room temperature. Yields calculated by relative integration of identifiable species seen by* ^{31}*P NMR. This is the same data shown in* ***Graph 2.5****, with yields shown for the first day only. The immediate conversion of* ***114b*** *into* ***129b****, and its conversion into* ***89b****, can be clearly seen.*

Hydrolysis of valine N-phosphordiamidate aminonitrile ***114e*** *in alkaline solution*



Characterisation data for the products observed (all 9:1 H_2O:D_2O):

129e: ^{1}H NMR (700 MHz) δ_H 3.40-3.37 (1H, m, (C2)-H, overlaps with **130e**), 2.01 (1H, dq, *J* = 13.2, 6.5 Hz, (C3)-H), 0.93 (3H, d, *J* = 6.7 Hz, (C4)-H_3, overlaps with **130e** but multiplets resolved), 0.83 (3H, d, *J* = 6.7 Hz, (C4′)-H_3). ^{13}C NMR (150 MHz) δ_C 181.5 (C1). ^{31}P NMR (284 MHz) δ_P12.05 (d, *J* = 9.5 Hz).

130e: ^{1}H NMR (700 MHz) δ_H 3.40-3.37 (1H, m, (C2)-H, overlaps with **129e**), 2.09-2.04 (1H, m, (C3)-H), 0.94 (3H, d, *J* = 7.0 Hz, (C4)-H_3, overlaps with **129e** but multiplets

resolved), 0.80 (d, *J* = 7.0 Hz, (C4′)-H_3). ^{13}C NMR (150 MHz) δ_C 182.7 (C1). ^{31}P NMR (284 MHz) δ_P 7.85 (d, *J* = 10.5 Hz).

82e: ^{1}H NMR (600 MHz) δ_H 3.20 (1H, dd, *J* = 10.6, 5.1 Hz, (C2)-H), 1.76 (1H, dq, *J* = 13.6, 6.9 Hz, (C3)-H), 0.80 (3H, d, *J* = 6.7 Hz, (C4)-H_3 – overlapping with (C4')-H_3 to give apparent triplet), 0.79 (3H, d, *J* = 6.6 Hz, (C4′)-H_3 – overlapping with (C4)-H_3 to give apparent triplet). ^{13}C NMR (150 MHz) δ_C 183.8 (d, *J* = 3.3 Hz, C1), 63.6 (d, *J* = 9.4 Hz, C2), 33.3 (dd, *J* = 13.8, 7.2 Hz, C3), 19.1 (C4 and C4′ overlapping). ^{31}P NMR (284 MHz) δ_P 9.04 (*J* = 10.5 Hz).

Val: ^{1}H NMR (600 MHz) δ 2.95 (1H, d, *J* = 5.2 Hz, (C2)-H), 1.81 (1H, q, *J* = 6.2 Hz, (C3)-H), 0.83 (3H, d, *J* = 6.7 Hz, (C4)-H_3), 0.76 (3H, d, *J* = 6.9 Hz, (C4′)-H_3). ^{13}C NMR (150 MHz) δ_C 169.1 (C-1), 62.7 (C2), 32.5 (C3), 19.8 (C4), 17.5 (C4′).

127e: (700 MHz) δ 3.11 (1H,d ,*J* = 6.2 Hz, (C2)-H).

Time / d	114e	129e	130e	82e	DAP	MAP	Val	127e
0	75%	0%	0%	0%	7%	10%	0%	0%
1	62%	12%	0%	0%	10%	11%	0%	0%
2	48%	22%	1%	0%	12%	11%	0%	0%
3	42%	26%	2%	0%	14%	12%	0%	0%
4	32%	33%	3%	0%	16%	12%	0%	0%
5	26%	36%	4%	1%	17%	12%	1%	2%
6	21%	38%	5%	2%	18%	13%	2%	3%
7	16%	40%	7%	2%	18%	13%	3%	3%
9	11%	41%	10%	3%	20%	14%	2%	3%
11	8%	40%	12%	4%	20%	14%	3%	3%
14	4%	38%	15%	5%	21%	14%	3%	3%
19	2%	34%	20%	7%	21%	15%	4%	3%
23	1%	30%	22%	8%	21%	15%	5%	3%
29	0%	24%	27%	10%	21%	16%	6%	2%
47	0%	13%	33%	15%	21%	17%	8%	2%

Table 7.2.60: *basic hydrolysis of purified* ***114e*** *at room temperature. Yields of phosphorylated produces are calculated by relative integration of identifiable species seen by ^{31}P NMR; yields of* ***Val*** *and* ***127e*** *are calculated by relative integration to* ***82e****. These yields are summarised in* ***Graph 2.6.***

Time / d	114e	129e	130e	82e	DAP	MAP
0	77%	0%	0%	0%	15%	8%
1	0%	18%	21%	9%	41%	11%
2	0%	6%	24%	18%	40%	11%
3	0%	2%	23%	23%	40%	12%
4	0%	0%	17%	30%	40%	12%
5	0%	0%	11%	38%	39%	12%
7	0%	0%	6%	43%	40%	11%

Table 7.2.61: *basic hydrolysis of purified* ***114e*** *at 50 °C. Yields shown are calculated by relative integration of identifiable species seen by* ^{31}*P NMR.*

Time / d	114e	129e	130e	82e	DAP	MAP	Val
9	0%	0%	3%	46%	34%	18%	10%
12	0%	0%	1%	47%	34%	18%	10%
15	0%	0%	0%	47%	34%	19%	10%

Table 7.2.62: *basic hydrolysis of purified* ***114e*** *at 50 °C. Yields of phosphorylated products are calculated by relative integration of identifiable species seen by* ^{31}*P NMR; the yield of* ***Val*** *is calculated by relative integration to* ***82e*** *by* ^{1}*H NMR.*

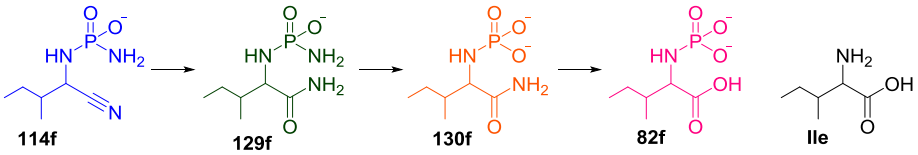
Time / d	114e	129e	130e	82e	DAP	MAP	Val	127e
0	12%	1%	0%	0%	71%	4%	0%	0%
1	0%	3%	3%	2%	74%	7%	1%	>1%
2	0%	1%	4%	4%	74%	7%	1%	>1%
3	0%	0%	3%	5%	74%	7%	1%	>1%
6	0%	0%	3%	6%	73%	7%	2%	>1%
8	0%	0%	2%	7%	73%	8%	2%	0%
11	0%	0%	1%	8%	73%	8%	2%	0%
14	0%	0%	1%	9%	73%	8%	2%	0%
18	0%	0%	0%	9%	73%	8%	3%	0%

Table 7.2.63: *basic hydrolysis of unpurified* ***114e*** *at 50 °C. Yields of phosphorylated produces are calculated by relative integration of identifiable species seen by* ^{31}*P NMR; yields of* ***Val*** *and* ***127e*** *are calculated by relative integration to* ***82e****.*

Time / d	82e (by ^{31}P)	DAP (by ^{31}P)	MAP (by ^{31}P)	Val (by ^{1}H)	82e (by ^{1}H)	Total (by integration of all peaks in methyl region)
18	7	72	11	20%	54%	100%

Table 7.2.64: *basic hydrolysis of purified* ***114e*** *at 50 °C, spiked with internal NMR standard. Yields of phosphorylated products are calculated by relative integration of identifiable species seen by* ^{31}*P NMR; the yields of all products are calculated by relative integration to internal standard by* ^{1}*H NMR.*

*Hydrolysis of isoleucine N-phosphordiamidate aminonitrile **114f** in alkaline solution*



Characterisation data for the products observed (all 9:1 H_2O:D_2O):

129f (two diastereomers): ^{1}H NMR (700 MHz) δ_H 3.57 (dd, *J* = 10.8, 4.3 Hz, (C2)-H), 3.45 (1H, dd, *J* = 10.7, 5.3 Hz, (C2)-H). ^{31}P NMR (284 MHz) δ_P 11.04 (d, *J* = 9.5 Hz), 10.96 (d, *J* = 8.6 Hz).

130f (two diastereomers): ^{1}H NMR (700 MHz) δ_H 3.55 (dd, *J* = 10.5, 3.6 Hz, (C2)-H), 3.46 (1H, dd, *J* = 10.9, 4.2 Hz, (C2)-H). ^{31}P NMR (284 MHz) δ_P 7.69 (d, *J* = 10.5 Hz), 7.63 (d, *J* = 10.5 Hz).

82f, diastereomer A (data assigned as far as possible): ^{1}H NMR (600 MHz) δ_H 3.30 (1H, dd, *J* = 10.9, 4.5 Hz, (C2)-H), 1.54-1.47 (1H, m, (C3)-H), 1.45-1.36 (1H, m, (C4)-H, overlapping with diastereomer B), 1.34-1.26 (1H, m, (C4)-H′, overlapping with diastereomer B), 0.84-0.79 (3H, overlapping with others, (C5)-H_3), 0.77 (3H, d, *J* = 6.9 Hz, (C3)-CH_3). ^{13}C NMR (150 MHz) δ_C 184.3 (d, *J* = 2.2 Hz, C1), 62.5 (br s, C2), 40.4 (C3). ^{31}P NMR (284 MHz) δ_P 9.45 (d, *J* = 10.5 Hz).

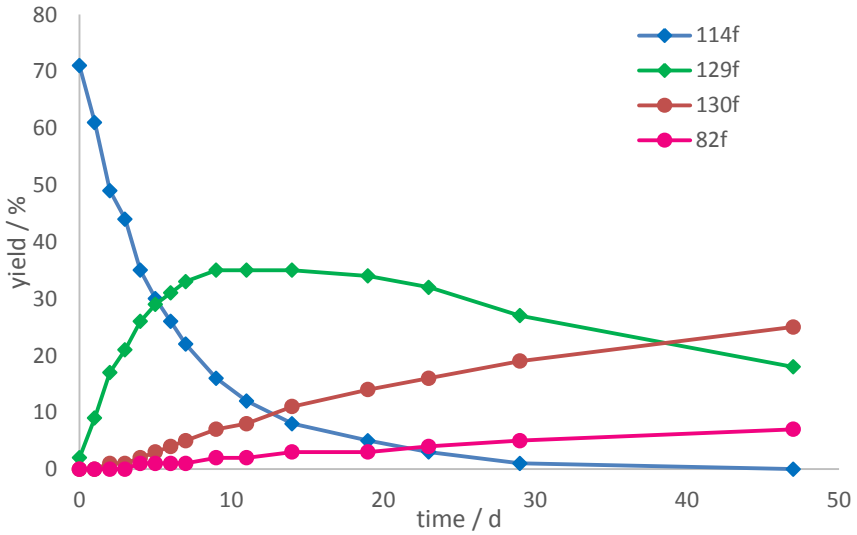
82f, diastereomer B (data assigned as far as possible): ^{1}H NMR (600 MHz) δ_H 3.27 (1H, dd, *J* = 10.6, 5.2 Hz), 2.12 (1H, dq, *J* = 13.6, 6.7 Hz, (C2)-H), 1.45-1.36 (1H, m, (C4)-H, overlapping with diastereomer A), 1.34-1.26 (1H, m, (C4)-H′, overlapping with diastereomer A), 0.96 (3H, d, *J* = 7 Hz, (C3)-CH_3), 0.84-0.79 (3H, overlapping with others, (C5)-H_3). ^{13}C NMR (150 MHz) δ_C 183.8 (d, *J* = 3.9 Hz, C1), 62.9 (d, *J* = 3.3 Hz, C2), 45.4 (C3), 18.2 ((C3)-CH_3). ^{31}P NMR (284 MHz) δ_P 9.23 (d, *J* = 10.5 Hz). HRMS (*m/z*): $[M+H^+]^-$ calcd for formula $C_6H_{12}NO_5PH^-$, 210.0531; found 210.0531.

Ile, diastereomer A (data assigned as far as possible): ^{1}H NMR (600 MHz) δ_H 3.15 (1H, d, *J* = 4.1 Hz, (C2)-H), 1.69-1.63 (1H, m, (C3)-H), 0.71 (3H, d, *J* = 6.9 Hz). ^{13}C NMR (150 MHz) δ_C 184.2 (C1), 59.9 (C2), 38.9 (C3).

Ile, diastereomer A (data assigned as far as possible): ^{1}H NMR (600 MHz) δ_H 3.01 (1H, d, *J* = 5.4 Hz, (C2)-H), 1.60-1.54 (1H, m, (C3)-H). ^{13}C NMR (150 MHz, 9:1 H_2O:D_2O) δ_C 183.9 (C1), 61.9 (C2), 39.4 (C3).

Time / d	114f	129f	130f	82f	DAP	MAP
0	71%	2%	0%	0%	13%	10%
1	61%	9%	0%	0%	15%	11%
2	49%	17%	1%	0%	18%	11%
3	44%	21%	1%	0%	19%	11%
4	35%	26%	2%	1%	21%	12%
5	30%	29%	3%	1%	22%	12%
6	26%	31%	4%	1%	23%	12%
7	22%	33%	5%	1%	24%	12%
9	16%	35%	7%	2%	25%	13%
11	12%	35%	8%	2%	26%	13%
14	8%	35%	11%	3%	27%	13%
19	5%	34%	14%	3%	28%	14%
23	3%	32%	16%	4%	30%	14%
29	1%	27%	19%	5%	29%	14%
47	<1%	18%	25%	7%	30%	16%

Table 7.2.65: *basic hydrolysis of purified* ***114f*** *at room temperature. Yields shown are calculated by relative integration of identifiable species seen by* ^{31}*P NMR. These results are shown in* ***Graph 7.2.5.***



Graph 7.2.5: *basic hydrolysis of purified* ***114f*** *at room temperature. Yields shown are calculated by relative integration of identifiable species seen by* ^{31}*P NMR.*

Time / d	114f	129f	130f	82f	DAP	MAP
0	75%	0%	0%	0%	11%	9%
1	2%	24%	9%	2%	47%	14%
2	0%	16%	17%	6%	46%	16%
3	0%	10%	20%	8%	46%	16%
4	0%	8%	21%	8%	46%	17%
5	0%	5%	22%	10%	46%	17%
7	0%	3%	23%	12%	45%	17%
8	0%	3%	22%	12%	45%	18%

Table 7.2.66: *basic hydrolysis of purified* ***114f*** *at 50 °C. Yields shown are calculated by relative integration of identifiable species seen by* ^{31}*P NMR.*

Time / d	114f	129f	130f	82f	DAP	MAP
9	0%	0%	7%	33%	38%	17%
12	0%	0%	4%	35%	39%	17%
16	0%	0%	1%	38%	39%	17%

Table 7.2.67: *basic hydrolysis of purified* ***114f*** *at 50 °C. Yields shown are calculated by relative integration of identifiable species seen by* ^{31}*P NMR.*

Time / d	114f	129f	130f	82f	DAP	MAP
0	15%	1%	0%	0%	62%	3%
1	0%	3%	4%	2%	66%	5%
2	0%	1%	6%	3%	66%	5%
3	0%	0%	5%	4%	67%	5%
6	0%	0%	4%	5%	67%	6%
8	0%	0%	3%	6%	67%	6%
11	0%	0%	2%	7%	66%	7%
14	0%	0%	2%	7%	66%	7%
18	0%	0%	1%	8%	67%	8%

Table 7.2.68: *basic hydrolysis of unpurified* ***114f*** *at 50 °C. Yields shown are calculated by relative integration of identifiable species seen by* ^{31}*P NMR.*

Time / d	130f	82f	Ile	All species present (by integration of all peaks in methyl region)
18	0.5%	36%	22%	82%

Table 7.2.69: *alkaline hydrolysis of purified* ***114f*** *at 50 °C, spiked with internal NMR standard: yields are calculated by relative integration to internal standard by* ^{1}*H NMR.*

7.2.5.3 Thiolysis of *N*-phosphordiamidate aminonitrile 114

General procedure 2.6: $NaSH.xH_2O$ (1.0 mmol) was dissolved in degassed water (100 µL) and adjusted to pH 9.0 with the addition of degassed 4 M HCl. This solution was then added to a degassed stock solution of either purified or unpurified *N*-phosphoroaminonitrile (0.05 mmol), diluted with water to a total volume of 500 µL and

stirred at room temperature. Aliquots (100 µL) were periodically removed, diluted with degassed D_2O (400 µL) and subjected to NMR analysis.

Thiolysis of valine N-phosphordiamidate aminonitrile ***114e***

Characterisation data for the products observed (all 9:1 H_2O:D_2O):

131e: ^{1}H NMR (600 MHz) δ_H 3.86 (1H, d, *J* = 10.9, 3.7 Hz, (C2)-H), 2.59-2.54 (1H, m, (C-3)-H), 1.12 (3H, d, J = 6.9 Hz, (C4)-H_3), 0.79 (3H, d, *J* = 7.0 Hz, (C4′)-H_3). ^{13}C NMR (150 MHz) δ_c 212.4 (C1), 68.6 (d, *J* = 16.0 Hz, C2), 33.3 (dd, *J* = 9.7, 6.9 Hz, C3), 20.7 (d, *J* = 4.9, C4), 15.6 (C4′). ^{31}P NMR (284 MHz) δ_P 7.43 (d, *J* = 10.5 Hz).

134f: ^{1}H (600 MHz) δ_H 3.47 (1H, d, *J* = 6.9 Hz, (C2)-H), 2.04 (1H, app sext, *J* = 6.9 Hz, (C3)-H), 1.00 (6H, d, *J* = 6.7 Hz, (C4)-H_3 and (C4′)-H_3 overlapping). ^{13}C NMR (150 MHz) δ_c 213.3 (C1), 67.2 (C2), 34.5 (C3), 18.0 (C4 and C4′ overlapping). LRMS m/z (CI+) 133.0 (100%, $[M+H^+]^+$), 187.1 (65%, $[M+3(NH_3)]^+$)

Time / d	114e	131e	DAP	MAP
0	76%	3%	16%	5%
1	63%	14%	17%	7%
4	36%	40%	18%	7%
6	13%	51%	19%	7%
8	11%	63%	18%	7%
11	2%	61%	19%	18%
15	0%	44%	20%	36%
22	0%	0%	21%	79%

Table 7.2.70: *thiolysis of purified valine N-phosphoroaminonitrile at room temperature. Yields shown are calculated by relative integration of identifiable species seen by ^{31}P NMR. ^{1}H NMR spectra shown in* ***Figure 7.2.8*** *and ^{31}P NMR spectra shown in* ***Figure 7.2.9.***

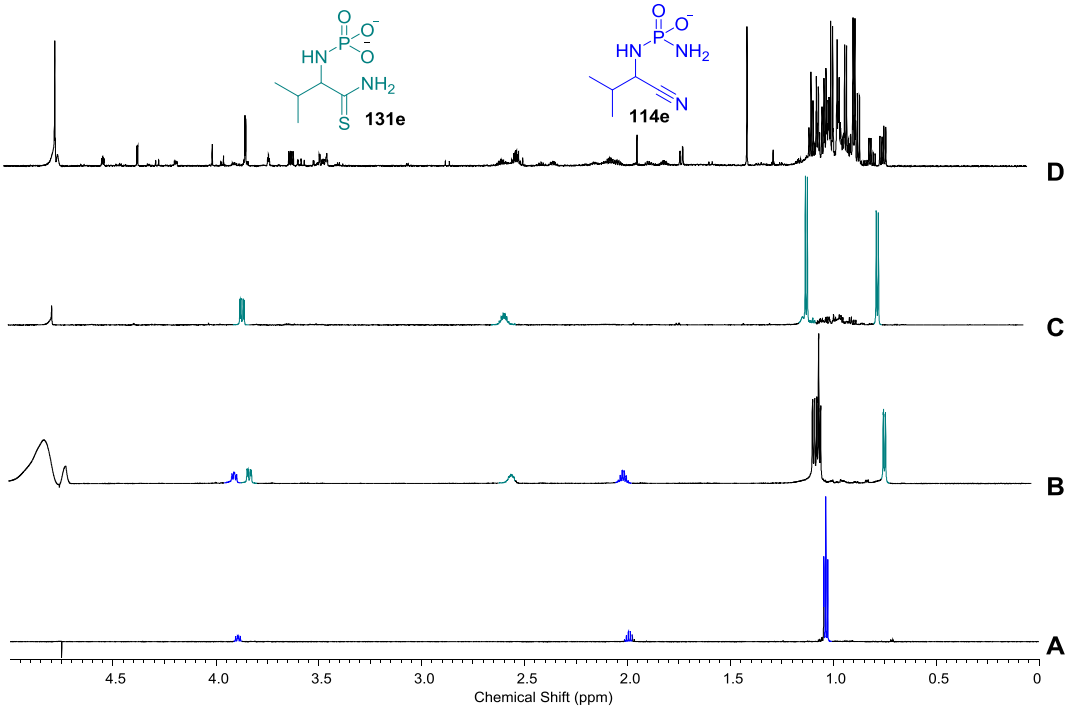


Figure 7.2.8: *Water-supressed ^{1}H NMR spectra (H_2O:D_2O 1:9, 0.00–5.00 ppm, 700 MHz) showing the reaction of* ***114e*** *(100mM) with NaSH at pH 9 at room temperature after:* ***A****) 30 mins;* ***B****) 4 d,* ***C****) 15 d, and* ***D****) 22 d.*

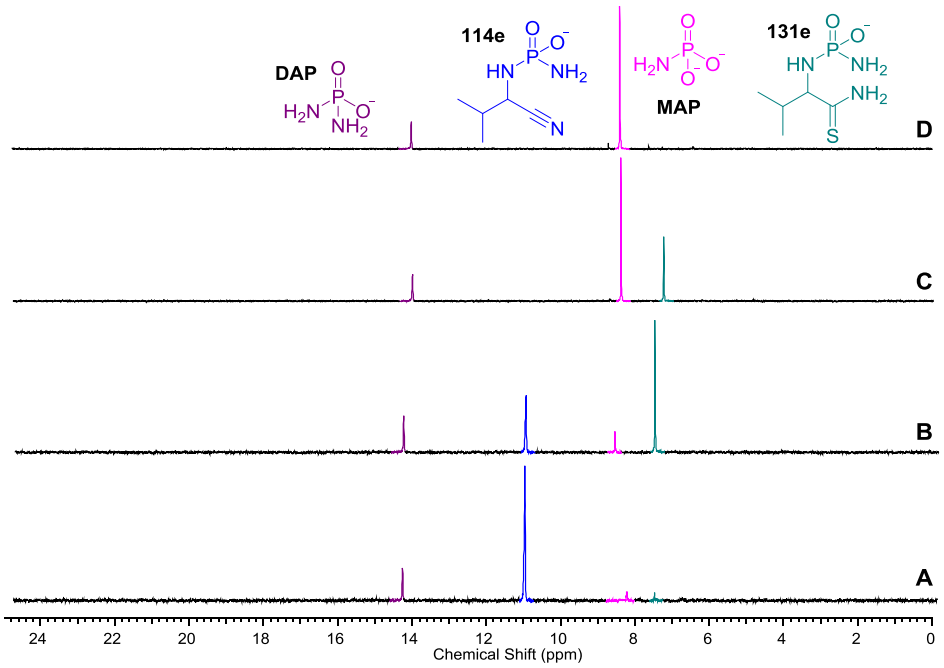


Figure 7.2.9: *31P NMR spectra (H_2O:D_2O 1:9, 0.0–25.0 ppm, 284 MHz) to show the reaction of* ***114e*** *(100mM) with NaSH at pH 9 at room temperature after:* ***A****) 30 mins;* ***B****) 4 d,* ***C****) 15 d, and* ***D****) 22 d.*

Time / d	114e	131e	DAP	MAP	134e
9	9%	54%	9%	20%	10%
12	5%	52%	10%	23%	13%
15	3%	51%	10%	25%	14%
16	2%	48%	11%	27%	16%

Table 7.2.71: *thiolysis of purified* ***114e*** *at room temperature. Yields of phosphorylated products shown are calculated by relative integration of identifiable species seen by* ^{31}*P NMR; the yield of unphosphorylated* ***134e*** *is calculated by relative integration to* ***114e*** *and* ***131e*** *by* ^{1}*H NMR.* ^{1}*H NMR spectra shown in* ***Figure 7.2.10.***

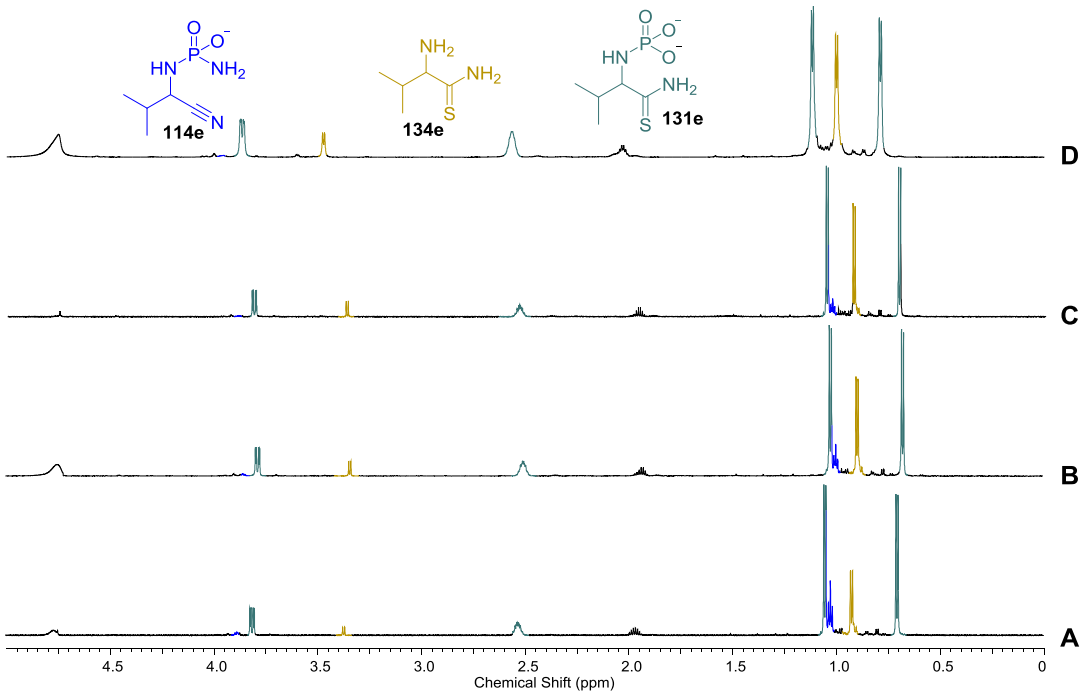


Figure 7.2.10: *Water-supressed* ^{1}*H NMR spectra (*H_2O*:*D_2O *1:9, 0.00–5.00 ppm, 700 MHz) to show the reaction of* ***114e*** *(100mM) with NaSH at pH 9 at room temperature after:* ***A****) 9d;* ***B****) 12 d,* ***C****) 15 d,* ***D****) 16 d.*

Time / d	114e	131e	DAP	MAP
0	14	0	69	0
1	10	3	71	3
2	7	5	71	4
3	6	7	70	4
6	3	9	71	4
8	2	10	71	4
11	1	11	74	4
14	0	10	71	6

Table 7.2.72: *thiolysis of unpurified* ***114e*** *at room temperature. Yields shown are calculated by relative integration of identifiable species seen by* ^{31}*P NMR.*

Time / d	134e (by ^{31}P)	DAP (by ^{31}P)	MAP (by ^{31}P)	134e (by ^{1}H)	131e (by ^{1}H)	Total (by integration of all peaks in methyl region)
18	7	72	7	27%	62%	97%

Table 7.2.73: *thiolysis of unpurified* ***114e*** *at room temperature, spiked with internal standard. Yields shown are calculated by relative integration of identifiable species seen by ^{31}P NMR, and by relative integration to internal standard.*

Thiolysis of isoleucine N-phosphordiamidate aminonitrile ***114f***

Characterisation data for the products observed (all 9:1 H_2O:D_2O):

131f, diastereomer A: ^{1}H NMR (600 MHz) δ_H 3.97 (1H, dd, *J* = 10.5, 3.0 Hz, (C-2)-H), 2.34-2.27 (1H, m, (C3)-H, overlaps with diastereomer B), 1.60-1.52 (1H, m, (C4)-H), 1.41-1.35 (1H, m, (C4)-H′), 1.01 (3H, t, *J* = 7.4 Hz, (C5)-H_3), 0.74 (3H, d, *J* = 7.0 Hz, (C3)-CH_3). ^{13}C NMR (150 MHz) δ_C 212.7 (d, *J* = 2.8 Hz, C1), 67.0 (C2), 40.2 (C3), 27.8 (C4), 12.7 ((C3)-CH_3 or C5), 12.6 ((C3)-CH_3 or C5). ^{31}P NMR (284 MHz) δ_P 7.32 (d, *J* = 10.5 Hz).

131f, diastereomer B: ^{1}H NMR (600 MHz) δ_H 3.88 (1H, dd, *J* = 11.1, 3.6 Hz, (C-2)-H), 2.34-2.27 (1H, m, (C3)-H, overlaps with diastereomer A), 1.27-1.22 (1H, m, (C4)-H), 1.09 (3H, d, *J* = 7.0 Hz, (C3)-CH_3), 0.98-0.94 (1H, m, (C4)-H′), 0.91 (3H, t, *J* = 7.4 Hz, (C5)-H_3). ^{13}C NMR (150 MHz) δ_C 212.3 (C1), 68.9 (C2), 40.2 (C3), 23.1 (C4), 17.1 (d, *J* = 2.8 Hz, (C3)-CH_3), 12.8 (C5). ^{31}P NMR (284 MHz) δ_P 7.42 (d, *J* = 11.4 Hz).

134f, diastereomer A (NMR data assigned as far as possible): ^{1}H NMR (600 MHz) δ_H 3.66 (1H, d, *J* = 5.5 Hz, (C2)-H), 1.95-1.91 (1H, m, (C3)-H). ^{13}C NMR (150 MHz) δ_C 213.7 (C1).

134f, diastereomer B (NMR data assigned as far as possible): ^{1}H NMR (600 MHz) δ_H 3.53 (1H, d, *J* = 7.2 Hz, (C2)-H), 1.84-1.78 (1H, m, (C3)-H). ^{13}C NMR (150 MHz, 9:1 H_2O:D_2O) δ_C 213.3 (C1). LRMS m/z (ESI+) 147.1 (100%, $[M+H^+]^+$)

Time / d	114f	131f	DAP	MAP
0	74%	0%	14%	5%
1	65%	8%	13%	7%
4	40%	25%	17%	9%
6	30%	33%	18%	10%
8	21%	40%	18%	12%
11	12%	45%	21%	12%
15	8%	45%	20%	14%

Table 7.2.74: *thiolysis of purified* ***114f*** *at room temperature. Yields shown are calculated by relative integration of identifiable species seen by ^{31}P NMR.*

Time / d	114f	131f	DAP	MAP
0	16%	0%	62%	0%
1	14%	1%	63%	2%
2	12%	3%	61%	2%
3	11%	4%	62%	2%
6	6%	8%	63%	2%
8	4%	10%	62%	2%
11	3%	12%	62%	2%
14	2%	12%	62%	3%
18	0%	12%	66%	3%

Table 7.2.75: *thiolysis of unpurified* ***114f*** *at room temperature. Yields shown are calculated by relative integration of identifiable species seen by ^{31}P NMR.*

Time / d	114f (by ^{1}H)	134f (by ^{1}H)	All species present (by integration of all peaks in methyl region)	114f (by ^{31}P)	131f (by ^{31}P)	DAP	MAP
18	61%	19%	82%	<1%	8%	65%	4%

Table 7.2.76: *thiolysis of unpurified* ***114f*** *at room temperature, spiked with internal NMR standard. Yields of phosphorylated products are calculated by relative integration of identifiable species seen by ^{31}P NMR; the yields of all products are calculated by relative integration to internal standard by ^{1}H NMR.*

Thiolysis of alanine N-phosphordiamidate aminonitrile ***114b***

Characterisation data for the products observed (all 9:1 H_2O:D_2O):

131b: ^{1}H NMR (600 MHz) δ_H 4.05-3.98 (1H, m, (C2)-H), 1.53 (3H, d, *J* = 7.2 Hz, (C3)-H_3). ^{13}C NMR (150 MHz) δ_C 213.7 (C1), 59.5 (d, *J* = 21.5 Hz, C2), 24.4 (C3). ^{31}P NMR (284 MHz) δ_P 7.50 (d, *J* = 10.5 Hz). LRMS m/z (ES-) 182.43 ($C_3H_7N_2O_3P$).

130b: ^{1}H NMR (600 MHz) δ_H 3.70-3.65 (1H, m, (C2)-H), 1.39 (1H, d, *J* = 7.0 Hz, (C3)-H). ^{13}C NMR (150 MHz) δ_C183.7 (d, *J* = 5.6 Hz, C1), 52.2 (dd, *J* = 10.4, 1.4 Hz, C2), 21.3 (C3). ^{31}P NMR (284 MHz) δ_P 8.00 (d, *J* = 10.5 Hz).

135a (unidentified): ^{31}P NMR (284 MHz, 9:1 H_2O:D_2O) δ_P 8.44 (d, *J* = 9.5 Hz).

135b (unidentified): ^{31}P NMR (284 MHz, 9:1 H_2O:D_2O) δ_P 6.59 (d, *J* = 10.5 Hz).

Time / d	114b	131b	130b	135a	135b	DAP	MAP
0	80%	0%	0%	0%	0%	20%	0%
1	38%	35%	0%	0%	0%	22%	5%
4	10%	54%	4%	0%	2%	23%	5%
6	2%	54%	10%	4%	3%	23%	6%
9	0%	39%	20%	8%	3%	21%	7%
12	0%	27%	30%	8%	2%	24%	8%

Table 7.2.77: *thiolysis of purified* ***114b*** *at room temperature. Yields shown are calculated by relative integration of species observed by ^{31}P NMR. These results are shown in* ***Graph 2.8.*** *^{1}H NMR spectra shown in* ***Figure 7.2.11.***

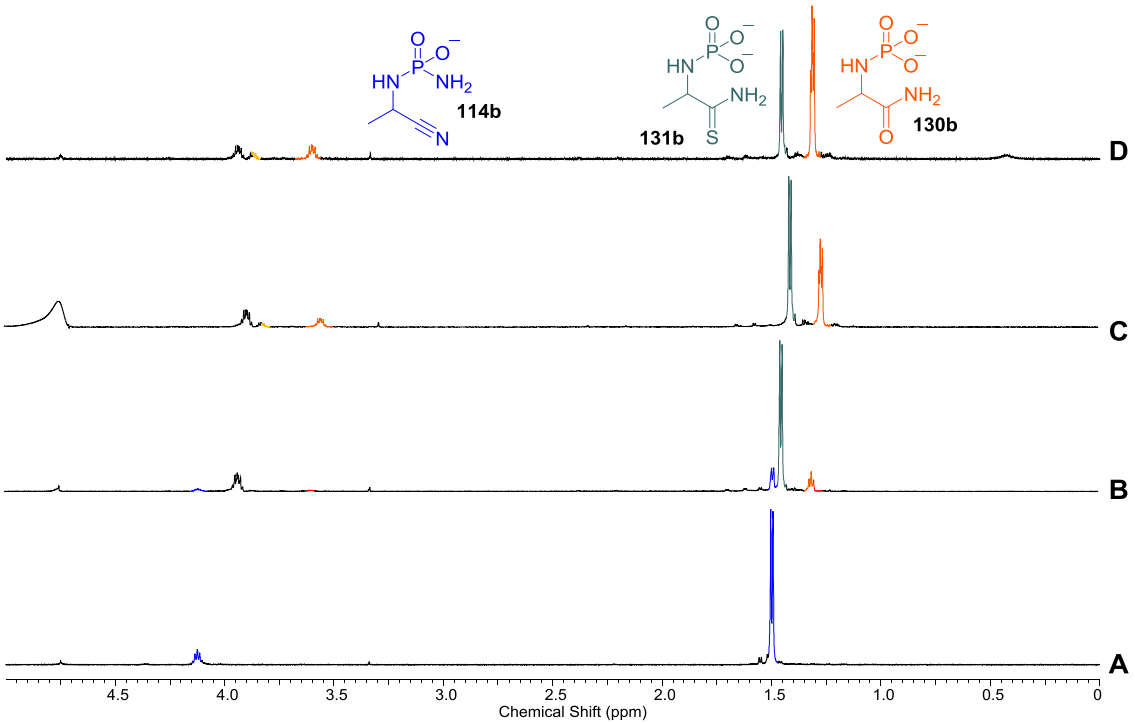


Figure A.6.40: *Water-supressed ^{1}H NMR spectra (H_2O:D_2O 1:9, 0.00–5.00 ppm, 700 MHz) showing the reaction of* ***114b*** *(100mM) with NaSH at pH 9 at room temperature after:* ***A****) 30 mins;* ***B****) 4 d,* ***C****) 9 d,* ***D****) 12 d.*

Time / d	114b	131b	130b	135a	135b	DAP	MAP
13	0%	37%	17%	3%	6%	23%	12%

Table 7.2.78: *thiolysis of purified* ***114b*** *at room temperature. Yields shown are calculated by relative integration of species observed by ^{31}P NMR.*

Time / d	114b	131b	130b	135a	135b	DAP	MAP
0	20%	0%	0%	0%	0%	75%	0%
1	10%	9%	0%	0%	0%	73%	2%
2	5%	12%	0%	0%	1%	73%	3%
3	3%	14%	0%	0%	2%	73%	3%
4	1%	13%	1%	0%	3%	74%	3%
5	1%	13%	1%	0%	3%	74%	3%
8	0%	10%	4%	0%	3%	73%	3%
11	0%	8%	7%	1%	2%	72%	4%
14	0%	5%	11%	1%	1%	72%	4%
18	0%	2%	12%	1%	1%	73%	5%
22	0%	1%	12%	1%	1%	74%	5%

Table 7.2.79: *thiolysis of unpurified* ***114b*** *at room temperature. Yields shown are calculated by relative integration of species observed by ^{31}P NMR. A second reaction on the same scale was prepared at the same time, and spiked with pentaerythritol and subjected to NMR analysis after 5 days. This indicated 72% conversion to* ***131b.*** *Integration of all methyl peaks indicated 95% of initial* ***114b*** *could be accounted for.*

7.3 Experimental section for borate as a phosphate analogue

7.3.1 Reactions of ancitabine

7.3.1.1 Ancitabine and boric acid

*Reaction of ancitabine ara-***63** *and boric acid*

To a 0.7M solution of H_3BO_3 in D_2O (2 ml, 1.4 mmol) at pH 6.2 was added ancitabine *ara*-**63** (80.0 mg, 0.31 mmol). The solution was stirred at 50 °C for 4 h, at which point solution was measured as pH 5.7. The reaction mixture was analysed by NMR. Conversion of 6% to arabinose-cytosine **21** was observed by ^{1}H NMR integral analysis. The presence of hydrolysis product **21** was confirmed by spiking with an authenticated sample. Arabinose cytosine **21**: ^{1}H NMR (600 MHz, D_2O) δ 7.85 (1H, d, *J* = 7.5 Hz, H-6), 6.22 (1H, d, *J* = 4.8 Hz, H-1′), 6.08 (1H, d, *J* = 7.5 Hz, H-5), 4.43 (1H, t, *J* = 4.4 Hz, H-2′), 4.14-4.16 (1H, m), 4.06-4.03 (1H, m), 3.95 (1H, dd, *J* = 3.3, 11.2 Hz, H-5′a), 3.87 (1H, dd, *J* = 12.3, 5.9 Hz, H-5′b). ^{13}C NMR (150 MHz, D_2O) δ 166.6 (C-4), 157.7 (C-2), 143.4 (C-6), 96.00 (C-5), 86.60 (C-1′), 83.82, 76.09, 75.95, 61.38 (C-5′).

General procedure 3.1 for the reaction of ancitabine *ara*-63 and boric acid at r.t.:

To a solution of H_3BO_3 in D_2O (570 - 750 mM, 750 µl-1 mL) was added ancitabine *ara*-**63**. The pH of the solution adjusted to the desired value, and placed in an NMR tube. ^{1}H NMR spectra were periodically acquired, and the contents of the NMR tube returned to an Eppendorf tube to allow measurement of the pH of the solution.

Tables showing the results of the reaction of ancitabine *ara*-63 and boric acid at r.t. Comparison of pH and quantity of boric acid on the formation of arabinose cytosine *ara*-**29** following general procedure 3.1. All percentages reported are proportional to the species integrated (the only species visible by ^{1}H NMR).

time / d	*ara*-**63** (% yield)	*ara*-**29** (% yield)	pH
0	100	0	6.1
1	100	0	
2	99	1	
3	97	3	
5	97	3	
6	96	4	6.3
8	96	4	
12	95	5	

Table 7.3.1: *Reaction of ancitabine ara-**63** (168 mM) with H_3BO_3 (3.4 eq), initial pH 6.1.*

General procedure 3.2 for the reaction of ancitabine *ara*-63 and boric acid at room temperature: to a 750 mM solution of H_3BO_3 in D_2O (750 µl, 1.0 mmol) was added ancitabine *ara*-**63**, the pH of the solution adjusted to the desired value, and placed in an NMR tube. ^{1}H NMR spectra were periodically acquired, and the contents of the NMR tube returned to an Eppendorf tube to allow measurement of the pH of the solution. The results of reactions carried out using general procedure 3.2 are shown in ***Tables 8.3.2-6*** and summarised in ***Table 3.1*****.**

time / d	*ara*-**63** (% yield)	*ara*-**29** (% yield)	pH
0	100	0	6.9
1	97	3	6.8
4	90	10	
5	87	13	6.8
11	54	46	

Table 7.3.2: *Reaction of ancitabine ara-**63** (116 mM) with 6.3 eq H_3BO_3, initial pH 6.9.*

time / d	*ara*-**63** (% yield)	*ara*-**29** (% yield)	pH
0	100	0	6.4
1	98	2	6.2
4	96	4	
5	95	5	6.1
7	85	15	
11	79	21	

Table 7.3.3: *Reaction of ancitabine ara-**63** (110 mM) with H_3BO_3, (6.8 eq) initial pH 6.4.*

time / d	*ara*-**63** (% yield)	*ara*-**29** (% yield)	pH
0	100	0	5.8
1	100	<1	5.6
4	98	2	
5	98	2	5.9
7	89	11	
11	88	12	

Table 7.3.4: *Reaction of ancitabine ara-**63** (115 mM) with H_3BO_3, (6.5 eq) initial pH 5.8.*

time / d	*ara*-**63** (% yield)	*ara*-**29** (% yield)	pH
0	100	0	5.6
1	100	0	5.3
4	99	1	
5	98	2	5.3
7	95	5	
11	94	6	

Table 7.3.5: *Reaction of ancitabine ara-**63** (119 mM) with H_3BO_3, (6.3 eq) initial pH 5.6.*

time / d	*ara*-**63** (% yield)	*ara*-**29** (% yield)	pH
0	100	0	4.8
1	100	0	4.6
4	100	0	
5	100	0	4.6
7	98	2	
11	97	3	

Table 7.3.5: Reaction of ancitabine ara-**63** (119 mM) with H_3BO_3 (6.3 eq), initial pH 5.6.

time / d	*ara*-**63** (% yield)	*ara*-**29** (% yield)	pH
0	100	0	6.3
3	100	0	
4	100	0	
10	98	2	

Table 7.3.6: Control reaction of ancitabine ara-**63** (84 mM) with no H_3BO_3, initial pH 6.3.

General procedure 3.3 for the reaction of ancitabine *ara*-63 and boric acid at 60 °C: to a solution of H_3BO_3 in D_2O (650 mM, 1 mL) was added ancitabine *ara*-**63** and additive. The pH of the solution adjusted to the desired value, and heated at 60 °C for 20 h. The pH of the reaction solution was then measured, and ^{1}H NMR spectra were acquired.

Control reactions: ancitabine *ara*-**63** was dissolved in in D_2O (1 mL). The pH of the solution adjusted to the desired value, and heated at 60 °C. The pH of the reaction solution was then measured, and ^{1}H NMR spectra were acquired.

Results shown in **Table 3.2**.

8.3.1.2 Reaction of ancitabine *ara*-63 and borax at r.t.

To a 175 mM solution of borax in D_2O (2 mL) was added ancitabine *ara*-**63** (69.27 mg, 0.26 mmol). The pH of the solution adjusted to 6.5, and placed in an NMR tube. ^{1}H NMR spectra were periodically acquired.

time / d	*ara*-**63** (% yield)	*ara*-**29** (% yield)
0	100	0
1	98	2
2	97	3
6	90	10
8	89	11
13	86	14
29	81	19

Table 7.3.7: *Reaction of ancitabine ara-**63**and borax at room temperature.*

8.3.1.3 Ancitabine and sodium perborate

Reaction of ancitabine *ara*-63 and sodium perborate at room temperature, initial pH 6.9

To a 130 mM solution of sodium perborate in D_2O (2 mL) was added ancitabine *ara*-**63** (23.94 mg, 0.09 mmol) and the pH of the solution adjusted to pH 6.9. A portion of the solution (500 µL) was placed in an NMR tube and ^{1}H NMR spectra periodically acquired; the remainder remained in an Eppendorf tube and the pH of the solution was periodically measured.

time / d	*ara*-**63** (% yield)	*ara*-**29** (% yield)	*ara*-**62** (% yield)	*ara*-**64** (% yield)	Unknowns (% yield)	pH
0	100	0	0	0	0	6.9
1	93	6	<1	<1	<1	4.2
2	84	11	<1	<1	5	-
5	87	14	1	<1	1	-
14	82	14	2	<1	2	3.8

Table 7.3.8: *Reaction of ara-**63** and sodium perborate at room temperature, initial pH 6.9.*

Reaction of ancitabine *ara*-63 and sodium perborate at 50 °C, initial pH 6.9

To a 130 mM solution of sodium perborate in D_2O (2 mL) was added ancitabine *ara*-**63** (19.01 mg, 0.07 mmol) and the pH of the solution adjusted to pH 6.9. A portion of the solution (500 µL) was placed in an NMR tube and the remainder was kept in an Eppendorf tube. ^{1}H NMR spectra were periodically acquired, and the pH of the solution was periodically measured.

time at 50 °C / h	*ara*-**63** (% yield)	*ara*-**29** (% yield)	*ara*-**62** (% yield)	*ara*-**64** (% yield)	Unknowns (% yield)	pH
0	100	0	0	0	0	6.9
4	92	<1	7	<1	<1	3.7
22	87	2	10	<1	<1	3.6
44	80	4	11	1	4	3.5

Table 7.3.9: *Reaction of ancitabine ara-**63** and sodium perborate at 50 °C, initial pH 6.9.*

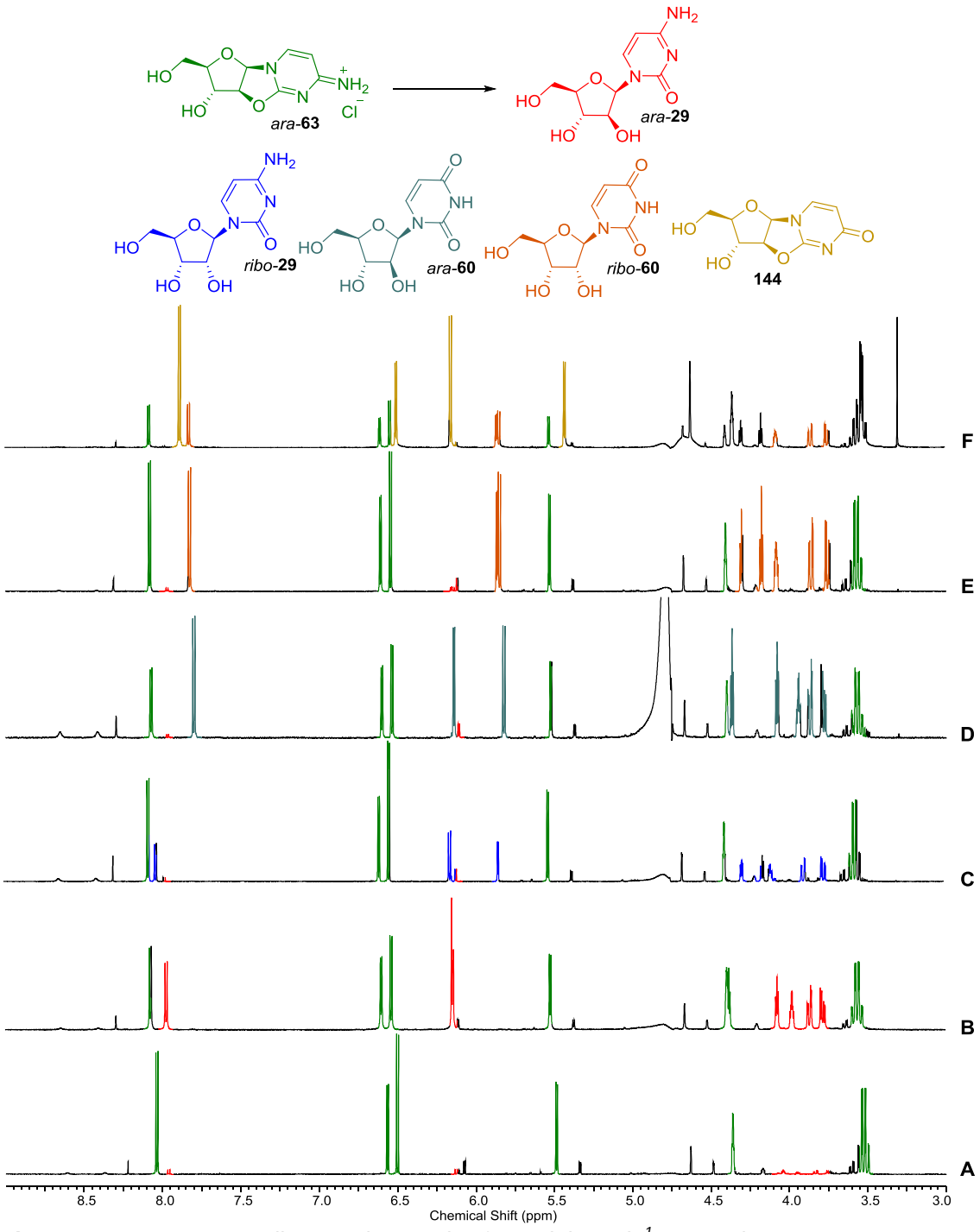


Figure 7.3.1: *Attempts to discover the species formed through ^{1}H NMR (600 MHz, 3.0-9.0 ppm, D_2O) spiking studies.* ***A)*** *^{1}H NMR spectra of the reaction of ancitabine ara-**63** (35 mM) with sodium perborate (130 mM) after 44 h at 50 °C. The reaction mixture was divided and each portion spiked with a different nucleoside.* ***B)*** *Positive spike with ara-**29**;* ***C)*** *negative spike with cytidine ribo-**31** ;* ***D)*** *negative spike with arabinose uracil ara-**60**;* ***E)*** *negative spike with uridine ribo-**60**, and this sample then used for* ***F)*** *negative spike with 2',2-cyclouridine **144**.*

General procedure 3.4 for the reaction of ancitabine *ara*-63 and sodium perborate at 60°C:

To a 140 mM solution of sodium perborate in D_2O (1 mL) was added ancitabine *ara*-**63** and additive. The pH of the solution adjusted to the desired value, and heated at 60 °C for 22 h. The pH of the reaction solution was then measured, and ^1H NMR spectra were acquired. Results shown in *Table 3.2.*

General procedure 3.5 for the reaction of ancitabine *ara*-63 and sodium perborate in H_2O

To a 140mM solution of $NaBO_3 \cdot 4H_2O$ in H_2O (7 ml, pH 6.8) was added ancitabine *ara*-**63** (30.5 mg, 0.12 mmol). Portions of this stock solution (1.6 mL) were adjusted to the desired pH and stirred at room temperature or 50 °C for 20 h. The pH of the solutions were then measured, and the reaction mixtures analysed by NMR. Conversion to arabinose-cytosine *ara*-**29**, arabinose amino-oxazoline *ara*-**62**, and arabinose oxazolidinone *ara*-**64** were confirmed by spiking with authenticated samples. Yields shown in ***Table 7.3.10***. Following ^1H NMR data assigned as far as possible:

ara-**29**: ^1H NMR (300 MHz, D_2O) δ 7.85 (1H, d, *J* = 7.8 Hz, H-6), 6.21 – 6.18 (1H, m, H-1′), 6.10-6.06 (1H, m, H-5),

ara-**62**: ^1H NMR (300 MHz, D_2O) δ 6.15 (1H, d, *J* = 5.5 Hz), 5.41 (1H, d, *J* = 5.5 Hz), 4.58 - 4.55 (1H, m), 4.27 – 4.23 (1 H, m).

ara-**64**: ^1H NMR (300 MHz, D_2O) δ 5.89 (1H, d, *J* = 5.8 Hz), 5.57 (1H, d, *J* = 6.0 Hz).

	ara-**63** (mmol)	$NaBO_3$ (eq.)	initial pH	Temp. (°C)	final pH	% yield of observed species (by relative integration of ^1H NMR spectra)				
						ara-**63**	*ara*-**29**	*ara*-**62**	*ara*-**64**	unknown
1	0.03	5.7	6.8	60	3.5	74	0	17	9	<1
2	0.03	5.7	7.7	60	4.8	9	<1	66	8	17
3	0.03	5.7	6.8	rt	4.3	88	0	12	0	<1
4	0.03	5.7	7.7	rt	6	29	<1	50	0	21

Table 7.3.10: *Reaction of ancitabine ara-**63** and sodium perborate in H_2O*

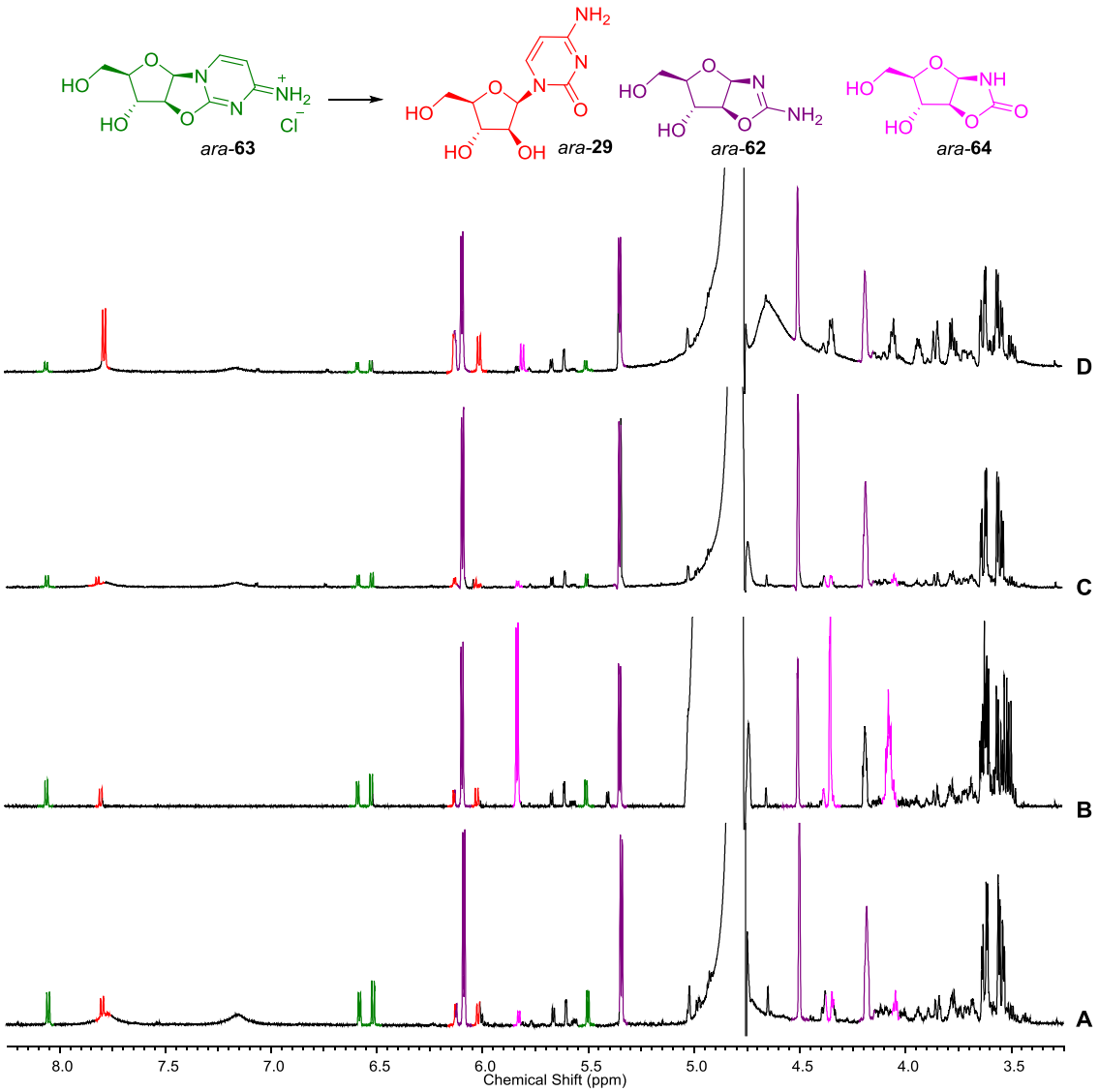


Figure 7.3.2: *Determining the species formed through ^{1}H NMR spiking studies.* ***A)*** *^{1}H NMR spectra of the reaction of ara-**63** (17 mM) with sodium perborate (140 mM) after 20 h at 60 °C. The reaction mixture was divided and each portion spiked with a different nucleoside.* ***B)*** *Positive spike with ara-**64**;* ***C)*** *positive spike with ara-**62**;* ***D)*** *positive spike with ara-**29**.*

8.3.2 Reactions of 2′,8-anhydroguanosine 148[1]

NMR data assigned as far as possible for the species observed:

2′,8-anhydroguanosine **148**: ^{1}H NMR (600 MHz, D_2O) δ 6.48 (1H, d, *J* = 5.5 Hz, H-1′), 5.71 (1H, d, *J* = 5.5 Hz, H-2′), 4.57 (1H, br s, H-3′), 4.30 – 4.26 (1H, m, H-4′), 3.45 (1H, ABX, *J* = 12.6, 4.6 Hz, H-5′a), 3.38 (1H, ABX, *J* = 12.8, 5.9 Hz, H-5′b). ^{13}C NMR (150 MHz, D_2O) δ 98.6 (C-2′), 86.2 (C-1′), 61.6 (C-5′).

[1] 2′,8-anhydroguanosine **148** prepared by Dr A. Nikmal.

5′,8-anhydro-arabinoguanosine *ara*-**149**: ^{1}H NMR (600 MHz, D_2O) δ 6.29 (1H, d, *J* = 6.5 Hz, H-1′), 4.56 (1H, H-5′a, overlaps with other peaks), 4.47 (1H, br s), 4.42 (1H, br s), 4.20 (1H, d, *J* = 12.2 Hz, H-5′b). ^{13}C NMR (150 MHz, D_2O) δ 87.5 (to 4.47), 84.8 (C-1′, 78.5 (to 4.42), 75.5 (C-5′).

2′,3′-epoxy-8-oxo-guanosine **150**: ^{1}H NMR (600 MHz, D_2O) δ 5.74 (1H, s, H-1′), 4.96 (1H, d, *J* = 2.6 Hz, H-2′), 4.31 (1H, app t, *J* = 5.1 Hz, H-4′), 3.68 (1H, ABX, *J* = 12.4, 5.4 Hz, H-5′a), 3.62 (1H, ABX, *J* = 12.2, 5.8 Hz, H-5′b). ^{13}C NMR (150 MHz, D_2O) δ 82.8 (C-1′), 60.3 (C-5′), 59.5 (C-2′).

8.3.2.1 2′,8-anhydro guanosine and boric acid

General procedure 3.6 for the reaction of 2′,8-anhydroguanosine 148 and boric acid: 2′,8-anhydro guanosine **148** (1.9 mg, 6.7 µmol) was added to a 430 mM solution of $B(OH)_3$ in D_2O (3 mL) and the pH adjusted from 4.79 to 6.45. The solution was split between two Eppendorf tubes: one was kept at room temperature, and the other at 60 °C. After 2 days, 0.6 mL of solution was taken from the room temperature sample, the pH adjusted to 8.2, and kept at 60 °C.

	% yield of nucleoside species (by relative integration of 1H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	6.5
1	100%	0%	0%	6.6
2	100%	0%	0%	6.6
3	100%	0%	0%	6.6
4	100%	0%	0%	6.6
7	100%	0%	0%	6.5

Table 7.3.11: *Reaction of 2.2 mM 2′,8-anhydroguanosine **148** with H_3BO_3 (183 eq) at room temperature, initial pH 6.5. (An aliquot was taken after 2 days, adjusted to pH 8.2 and heated: results shown in **Table 7.3.12**)*

	% yield of nucleoside species (by relative integration of 1H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	8.2
1	76%	14%	10%	8.2
2	64%	18%	18%	8.1
5	40%	35%	25%	8.2
8	24%	41%	35%	8.2

Table 7.3.12: *Reaction of 2.2 mM 2′,8-anhydroguanosine **148** with H_3BO_3 (183 eq) at 60°C, initial pH 8.2. (An aliquot of the reaction at room temperature, initial pH 6.5 was taken after 2 days, adjusted to pH 8.2 and heated. Reaction time from pH adjustment and heating.)*

	% yield of nucleoside species (by relative integration of ^{1}H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	6.5
1	95%	5%	0%	6.6
2	91%	9%	0%	6.6
3	84%	16%	0%	6.6
4	77%	23%	0%	6.6
7	58%	42%	0%	6.6

Table 7.3.13: *Reaction of 2.2 mM 2',8-anhydroguanosine* ***148*** *with H_3BO_3 (183 eq) at 60 °C, initial pH 6.5.*

General procedure 3.7 for the reaction of 2',8-anhydro guanosine 148 and boric acid: 2′,8-anhydro guanosine **148** (6.6 mg, 22.8 µmol) was added to a 620 mM solution of $B(OH)_3$ in D_2O (7 mL) and the pH adjusted from 4.5 to 7.3. The solution was split between seven Eppendorf tubes and the pH of each solution adjusted. All were stirred at room temperature for 9 days, and then stirred at 40 °C.

	% yield of nucleoside species (by relative integration of 1H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	7.1
4	100%	0%	0%	7.5
9	100%	0%	0%	7.5
11	100%	0%	0%	7.5
14	91%	9%	0%	7.5
16	90%	10%	0%	7.5

Table 7.3.14: *Reaction of 3.3 mM 2',8-anhydroguanosine* ***148*** *with H_3BO_3 (207 eq) at room temperature and then at 40 °C, initial pH 7.1.*

	% yield of nucleoside species (by relative integration of 1H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	7.5
4	100%	0%	0%	7.5
9	100%	0%	0%	7.5
11	100%	0%	0%	7.5
14	92%	8%	0%	7.5
16	90%	10%	0%	7.5

Table 7.3.15: *Reaction of 3.3 mM 2',8-anhydroguanosine* ***148*** *with H_3BO_3 (207 eq) at room temperature and then at 40 °C, initial pH 7.5.*

	% yield of nucleoside species (by relative integration of 1H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	7.9
4	100%	0%	0%	7.9
9	100%	0%	0%	7.8
11	100%	0%	0%	7.9
14	92%	8%	0%	7.8
16	87%	11%	3%	7.9

Table 7.3.16: *Reaction of 3.3 mM 2',8-anhydroguanosine* ***148*** *with H_3BO_3 (207 eq) at room temperature and then at 40 °C, initial pH 7.9.*

	% yield of nucleoside species (by relative integration of 1H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	8.6
4	100%	0%	0%	8.5
9	100%	0%	0%	8.5
11	89%	11%	0%	8.5
14	82%	14%	5%	8.5
16	83%	17%	11%	8.5

Table 7.3.17: *Reaction of 3.3 mM 2',8-anhydroguanosine* ***148*** *with H_3BO_3 (207 eq) at room temperature and then at 40 °C, initial pH 8.6.*

	% yield of nucleoside species (by relative integration of 1H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	9.0
4	100%	0%	0%	9.0
9	89%	11%	0%	9.0
11	69%	19%	11%	9.0
14	59%	25%	16%	9.1
16	60%	21%	19%	9.1

Table 7.3.18: *Reaction of 3.3 mM 2',8-anhydroguanosine* ***148*** *with H_3BO_3 (207 eq) at room temperature and then at 40 °C, initial pH 9.0.*

	% yield of nucleoside species (by relative integration of 1H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	9.6
4	100%	0%	0%	9.5
9	85%	15%	0%	9.5
11	52%	25%	23%	9.5
14	41%	29%	30%	9.5
16	47%	26%	27%	9.5

Table 7.3.19: *Reaction of 3.3 mM 2′,8-anhydroguanosine* ***148*** *with H_3BO_3 (207 eq) at room temperature and then at 40 °C, initial pH 9.6.*

	% yield of nucleoside species (by relative integration of 1H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	10.0
4	98%	0%	2%	9.8
9	69%	13%	18%	9.8
11	49%	26%	25%	9.8
14	35%	30%	35%	9.9
16	24%	27%	48%	9.9

Table 7.3.20: *Reaction of 3.3 mM 2′,8-anhydroguanosine* ***148*** *with H_3BO_3 (207 eq) at room temperature and then at 40 °C, initial pH 10.0.*

7.3.2.2 2′,8-anhydroguanosine and sodium perborate.

Reaction of 2.8 mM 2′,8-anhydroguanosine 148 with $NaBO_3$ (18 eq) at room temperature, initial pH 6.5.

2′,8-anhydroguanosine **148** (2.4 mg, 8.5 µmol) was added to a 0.05 mM solution of $NaBO_3 \cdot 4H_2O$ in D_2O (3 mL) and adjusted from pH 10.3 to 6.5. After 2 days, 1 mL of solution was taken, the pH adjusted to 8.1, and kept at 60 °C.

	% yield of nucleoside species (by relative integration of 1H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	6.5
1	100%	0%	0%	6.7
2	100%	0%	0%	6.8
3	100%	0%	0%	6.9
4	100%	0%	0%	6.8
7	100%	0%	0%	6.9

Table 7.3.21: *Reaction of 2.8 mM 2′,8-anhydroguanosine* ***148*** *with $NaBO_3$ (18 eq) at room temperature, initial pH 6.5. (An aliquot was taken after 2 days, adjusted to pH 8.1 and heated: results shown in* ***Table 7.3.22****)*

	% yield of nucleoside species (by relative integration of 1H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	8.1
1	91%	9%	0%	8.2
2	82%	18%	0%	8.1
5	51%	30%	19%	8.1
8	41%	29%	30%	-

Table 7.3.22: *Reaction of with 2.8 mM 2′,8-anhydroguanosine* ***148*** *with $NaBO_3$ (18 eq) at 60°C, initial pH 8.1. (An aliquot of the reaction at room temperature, initial pH 6.5 was taken after 2 days, adjusted to pH 8.1 and heated.)*

Reaction of 3.3 mM 2′,8-anhydroguanosine 148 with $NaBO_3$ (23 eq.)

2′,8-anhydroguanosine **148** (6.7 mg, 23.4 µmol) was added to a 80 mM solution of $NaBO_3$ in D_2O (7 mL) and the pH adjusted from 10.7 to 8.9. The solution was split between seven Eppendorf tubes and the pH of each solution adjusted. All were stirred at room temperature for 9 days, and then stirred at 40 °C.

	% yield of nucleoside species (by relative integration of 1H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	7.1
4	100%	0%	0%	7.1
9	100%	0%	0%	7.1
11	100%	0%	0%	7.2
14	100%	0%	0%	7.2
16	94%	6%	0%	7.3

Table 7.3.23: *Reaction of with 3.3 mM 2',8-anhydroguanosine* ***148*** *with $NaBO_3$ (23 eq) at room temperature and then at 40 °C, initial pH 7.1.*

	% yield of nucleoside species (by relative integration of 1H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	7.5
4	100%	0%	0%	7.5
9	100%	0%	0%	7.5
11	100%	0%	0%	7.5
14	100%	0%	0%	7.6
16	92%	8%	0%	7.6

Table 7.3.24: *Reaction of with 3.3 mM 2',8-anhydroguanosine* ***148*** *with $NaBO_3$ (23 eq) at room temperature and then at 40 °C, initial pH 7.5.*

	% yield of nucleoside species (by relative integration of 1H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	8.1
4	100%	0%	0%	8.1
9	100%	0%	0%	8.2
11	100%	0%	0%	8.3
14	100%	0%	0%	8.4
16	89%	11%	0%	8.5

Table 7.3.25: *Reaction of with 3.3 mM 2',8-anhydroguanosine* ***148*** *with $NaBO_3$ (23 eq) at room temperature and then at 40 °C, initial pH 8.1.*

	% yield of nucleoside species (by relative integration of 1H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	8.5
4	100%	0%	0%	8.4
9	100%	0%	0%	8.5
11	85%	11%	4%	8.6
14	77%	16%	7%	8.7
16	74%	16%	9%	8.8

Table 7.3.26: *Reaction of with 3.3 mM 2',8-anhydroguanosine* ***148*** *with $NaBO_3$ (23 eq) at room temperature and then at 40 °C, initial pH 8.5.*

	% yield of nucleoside species (by relative integration of 1H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	9.2
4	100%	0%	0%	9.1
9	81%	10%	8%	9.2
11	65%	20%	15%	9.3
14	55%	25%	21%	9.3
16	56%	26%	19%	9.4

Table 7.3.27: *Reaction of with 3.3 mM 2',8-anhydroguanosine* ***148*** *with $NaBO_3$ (23 eq) at room temperature and then at 40 °C, initial pH 9.2.*

	% yield of nucleoside species (by relative integration of 1H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	9.5
4	100%	0%	0%	9.5
9	76%	11%	13%	9.6
11	56%	21%	23%	9.6
14	41%	28%	31%	9.7
16	37%	29%	34%	9.7

Table 7.3.28: *Reaction of with 3.3 mM 2',8-anhydroguanosine* ***148*** *with $NaBO_3$ (23 eq) at room temperature and then at 40 °C, initial pH 9.5.*

	% yield of nucleoside species (by relative integration of 1H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	10.0
4	100%	0%	0%	9.9
9	68%	17%	15%	10.0
11	40%	26%	34%	10.1
14	34%	29%	37%	10.1
16	31%	31%	39%	10.1

Table 7.3.29: *Reaction of with 3.3 mM 2',8-anhydroguanosine* ***148*** *with $NaBO_3$ (23 eq) at room temperature and then at 40 °C, initial pH 10.0.*

	% yield of nucleoside species (by relative integration of 1H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	7.0
1	62%	17%	21%	-
2	39%	25%	36%	-
3	22%	32%	44%	9.6
4	17%	33%	50%	-

Table 7.3.30: *Reaction of 32 mM 2',8-anhydroguanosine* ***148*** *with $NaBO_3$ (220 mM, 6.9 eq) at 50 °C, initial pH 7.0.*

	% yield of nucleoside species (by relative integration of 1H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	6.5
1	85%	12%	3%	-
2	77%	17%	6%	-
3	77%	17%	6%	7.9
4a	77%	17%	6%	7.9
4b	47%	22%	31%	9.7

Table 7.3.31: *Reaction of 32 mM 2',8-anhydroguanosine* ***148*** *with H_3BO_3 (240 mM, 7.5 eq) at 50 °C, initial pH 6.5. After heating at 50 °C for 3 d the reaction mixture was divided in two. One portion was adjusted to pH 9.7; the other portion was left unadjusted (pH 7.9). Both portions were heated for a further day.*

	% yield of nucleoside species (by relative integration of 1H NMR spectra)			
time / d	**148**	*ara*-**149**	**150**	pH
0	100%	0%	0%	7.0
1	86%	14%	0%	-
2	85%	15%	0%	-
3	84%	16%	0%	7.2
4a	84%	16%	0%	7.2
4b	52%	21%	27%	9.8

Table 7.3.32: *32 mM 2',8-anhydroguanosine **148** with no borate at 50 °C, initial pH 7.0. After heating at 50 °C for 3 d the reaction mixture was divided in two. One portion was adjusted to pH 9.8; the other portion was left unadjusted (pH 7.2). Both portions were heated for a further day.*

8.4 Experimental section for *O*-acetylated purines from glycosylation reactions

8.4.1 Synthesis of acetylated ribose derivatives

1,2,3,5-tetra-acetyl-ribose 169

Ribose *ribo*-**30** (2.03 g, 13.54 mmol) was stirred in pyridine (30 mL) at 0°C, and acetic anhydride (8 mL, 84.6 mmol) and DMAP (74.6 mg, 0.61 mmol) added. The temperature of the mixture was then raised and stirred at room temperature for 28 h. The mixture was then concentrated by rotary evaporation, dissolved in EtOAc (50 mL) and washed with water (3 x 50 mL) and brine (50 mL), dried ($MgSO_4$) and concentrated *in vacuo*. The crude product was purified by flash column chromatography (25% EtOAc in hexanes to 20% MeOH in EtOAc) to afford 1,2,3,5-tetra-acetyl-ribose **169** (3.06 g, 71%) as a viscous colourless oil. **169** was obtained as a mixture of α- and ß-furanosyl and pyranosyl isomers, each with a discreet H-1 peak: δ 6.02 (d, *J* = 4.7 Hz, 62%), 6.41 (d, *J* = 4.5, 3%) 6.15 (s, 12%), and 6.05 (d, *J* = 3.5 Hz, 24%). Full NMR data is provided for the major species (probably ß-furanosyl): ^{1}H NMR ($CDCl_3$, 600 MHz) δ 6.02 (1H, d, *J* = 4.7 Hz), 5.47 (1H, app t, *J* = 3.4 Hz, H-3), 5.15-5.13 (1H, m, H-4), 5.03 (1H, ddd, *J* = 4.6, 3.7, 0.6 Hz, H-2), 4.01 (1H, ABX, *J* = 12.4, 3.4 Hz, H-5a), 3.90 (1H, ABX, *J* = 12.4, 5.7 Hz, H-5b), 2.12 (3H, s), 2.094 (3H, s), 2.087 (3H, s), 2.08 (3H, s). ^{13}C NMR ($CDCl_3$, 150 MHz) δ 170.05, 169.9, 169.6, 168.9 (C-1 Ac), 91.00 (C-1), 67.4 (C-2), 66.2 (br, C-2 and C-3), 62.8 (C-5), 21.0, 20.9, 20.82, 20.79. Rf = 0.7 (1:1 EtOAc: Hexanes). This data is in good agreement with literature. [402]

1,2,3-Tris-acetyl-ribose-5-phosphate 170

Table 7.4.1, **entry 8:** 1,2,3-Tris-acetyl-ribose-5-phosphate **170** was isolated in 20% yield (314.4 mg, 0.79 mmol) as a mixture of α and ß isomers (249.7 mg as 1:9 α:ß; 64.7 mg as 1:4 α:ß) as a colourless oil. ^{1}H NMR (D_2O, 600 MHz) δ 6.39 (1H, d, *J* = 4.6 Hz, ß-H-1), 6.10 (1H, s, α H-1), 5.43 – 5.38 (2H, m, H-2 and H-3), 4.46 – 4.41 (1H, m, H-4), 3.99 (1H, ABX, *J* = 5.7, 3.2, 2.3, H-5a), 3.92-3.85 (1H, m, H-5b), 2.11 (s, 3H), 2.09 (s, 3H), 2.05 (s, 3H). ^{13}C NMR (D_2O, 150 MHz) δ 173.5, 173.3, 173.2, 119.6, 99.0 (C-1ß), 95.3 (C-1α), 81.5 (d, *J* = 8.3 Hz, C-4), 75.1 (C-3?), 71.2 (C-2?), 65.0 (d, *J* = 4.4 Hz, C-5), 21.2, 20.64, 20.60. ^{31}P NMR (D_2O, 284 MHz) δ 0.37. This data is in good agreement with literature.[403]

Entry	Ribose-5-phosphate *ribo*-**30**-5P	*N*-acetyl imidazole **168**	solvent	pH	Conversion to acetylated *ribo*-**30**-5P products after 2 h	Additional **168** added after 2 h (eq.)	Conversion to acetylated *ribo*-**30**-5P products after 4 h
1	22.4 mg, 0.097 mmol	108.27 mg, 0.98 mmol	H_2O	8.0	70%	-	-
2	22.91 mg, 0.10 mmol	110.48 mg, 1.00 mmol	H_2O	7.5	85%	-	-
3	22.41 mg, 0.097 mmol	115.77 mg, 1.05 mmol	H_2O	7.0	88%	5.6	90%
4	22.18 mg, 0.096 mmol	121.68 mg, 1.11 mmol	1:1 H_2O/D_2O	8.0	83%	-	-
5	22.2 mg, 0.096 mmol	117.04 mg, 1.06 mmol	1:1 H_2O/D_2O	7.5	90%	2.9	90%
6	22.17 mg, 0.096 mmol	115.98 mg, 1.05 mmol	1:1 H_2O/D_2O	7.0	92%	-	-
7	225.3 mg, 0.98 mmol	1.23 g, 11.18 mmol	1:1 H_2O/D_2O	7.0	92%[a]	-	-
8	888.4 mg, 3.86 mmol	4.82 g, 43.7 mmol	3:1 H_2O/D_2O	7.0	92%[b]	-	-

Table 7.4.1*: Formation of 1,2,3-tris-acetyl-ribose-5-phosphate* ***170*** *from ribose-5-phosphate ribo-****30****-5P with N-acetyl imidazole* ***168.*** *All reactions carried out at 100 mM with respect to ribo-****30****-5P with 10 eq.* ***168*** *in 1 mL solvent except entry 7 (10 mL) and entry 8 (40 mL). Adjusted to desired pH with 4 M NaOH and readjusted when additional* ***168*** *added. All pH values are uncorrected. All 0.1 mmol ribo-****30****-5P except entry 7 (1 mmol) and entry 8 (3.9 mmol). The peaks for the various partially acetylated products overlap with those for* ***170*** *(the desired product), but are discrete from the unacetylated starting material ribo-****30****-5P. Conversion estimated by relative integration of H-4.*

[a] *crude product was purified by flash column chromatography (10% to 100% MeOH in EtOAc), then reverse-phase C18 flash column chromatography to afford 1,2,3-tris-acetyl-ribose-5-phosphate* ***170*** *in 3% yield (10.8 mg, 0.027 mmol).*

[b] *crude product purified by flash column chromatography, then reverse-phase C18 flash column chromatography.*

2,3,5-tri-acetyl-ribose

With hydrazine: 1,2,3,5-Tetra-acetyl-ribose **169** (1.27 g, 3.98 mmol) was dissolved in DMF (16 mL) and anhydrous hydrazine (0.25 mL, 8.03 mmol) added. The mixture was stirred for 2 h, then EtOAc (16 mL) and brine (16 mL) added, and the layers separated. The aqueous layer was extracted with EtOAc (3 x 16 mL), the combined organic layers washed with brine (16 mL) and dried ($MgSO_4$), and concentrated *in vacuo*. The crude product was purified by flash column chromatography (hexanes:EtOAc 3:1 to EtOAc:MeOH 9:1) to afford **171** contaminated with DMF and other species (261.8 mg, 0.95 mmol, 24%).

1,2,3,5-Tetra-acetyl-ribose (514.8 mg, 1.62 mmol) was dissolved in DMF (8 mL) and anhydrous hydrazine (0.13 mL, 4.17 mmol) added. The mixture was stirred at 50 °C for 2 h, then EtOAc (8 mL) and brine (8 mL) added, and the layers separated. The aqueous layer was extracted with EtOAc (5 x 16 mL), dried ($MgSO_4$), and concentrated *in vacuo*. The crude product was purified by flash column chromatography (hexanes:EtOAc 3:1 to EtOAc:MeOH 9:1) to afford **171** contaminated with other species (33.4 mg, 0.12 mmol, 7%).

General procedure 4.1 for bromination of 1,2,3-tris-acetyl-ribose-5-phosphate **170.**

1,2,3,5-Tetra-acetyl-ribose **170** was dissolved in CH_2Cl_2 (1 mL or 5 mL) and stirred at 0°C. HBr in AcOH (45% w/v) was added and the mixture stirred at 0°C. The reaction mixture was then diluted with water (5 or 25 mL) and CH_2Cl_2 (5 or 25 mL), separated, and the organic layer washed with water (2 x 5or 25 mL), dried ($MgSO_4$), and concentrated *in vacuo*. The residue was dissolved in $CDCl_3$ and subjected to ^{1}H NMR analysis. Results shown in ***Table 7.4.2.***

Entry	**170**	HBr in AcOH, 45% w/v	Reaction time / h	**169**	ß-**173**	α-**173**	**171**
1	107.8 mg, 0.34 mmol	0.3 mL, 1.7 mmol	0.5	19%	15%	35%	32%
2	103.6 mg, 0.33 mmol	0.3 mL, 1.7 mmol	1	6%	17%	42%	34%
3	106.6 mg, 0.33 mmol	0.3 mL, 1.7 mmol	1.5	1%	17%	48%	33%
4	516.8 mg, 1.62 mmol	1.5 mL, 8.5 mmol	1.5	22%	21%	43%	14%

Table 7.4.2. *Investigations of the bromination of 1,2,3-tris-acetyl-ribose-5-phosphate* ***170.*** *All reactions carried out using 1,2,3,5-tetra-O-acetyl ribose (340 mM) and HBr in AcOH (45% w/v, 5 eq.) in DCM. All on 0.3 mmol scale in 1 mL CH_2Cl_2 except entry 4 (1.6 mmol, 5 mL).*

The crude colourless oil from ***Table 7.4.2****, entry 4* was used in the next step: zinc (399.4 mg, 6.1 mmol) was stirred in a mixture of water (10 mL) and AcOH (10 mL) at 0 °C. Crude **173** was dissolved in Et_2O (15 mL) and added slowly to the zinc mixture over a period of 15 mins. The reaction mixture was stirred at room temperature for 20 h, then filtered, diluted with CH_2Cl_2 (15 mL), and washed with water (3 x 10 mL), saturated $NaHCO_3$ (3 x 10 mL) and brine (1 x 10 mL). The combined aqueous layers were extracted with DCM (3 x 50 mL), and the combined organic layers dried ($MgSO_4$) and concentrated *in vacuo.* ^{1}H NMR analysis revealed that the **173** had converted to **171**.

Preparative synthesis: 1,2,3,5-Tetra-acetyl-ribose **170** (534.6 mg, 1.68 mmol) was dissolved in CH_2Cl_2 (5 mL) and stirred at 0°C. HBr in AcOH (45% w/v; 1.5 mL) was added and the mixture stirred at 0°C for 1.5 h. Saturated aqueous $NaHCO_3$ .was then added and the resulting mixture stirred vigorously for a further 1.5 h. The mixture was then diluted with CH_2Cl_2 (30 mL), the layers separated, and the organic layer washed with brine (30 mL), dried ($MgSO_4$) and concentrated *in vacuo*. The crude product was then purified by flash column chromatography (25% EtOAc in hexane to 100% EtOAc to 10% MeOH in EtOAc) to afford 2,3,5-tri-acetyl-ribose (143.8 mg, 31%) as a colourless oil. Two products were observed by NMR analysis in a 82:18 ratio. IR (CM^{-1}) 1737, 1433, 1370, 1216, 1070-1020. HRMS (*m/z*) calcd for formula $[C_{11}H_{16}O_8Na]^+$ 299.0739; found 299.0739. Data in agreement with literature.[361]

Major product: ^{1}H NMR (600 MHz, $CDCl_3$) δ 5.51 (1H, apparent t, *J* = 3.2 Hz, H-2, obscures minor product H-2), 5.12-5.08 (2H, m, H-1 and H-4), 4.94 (1H, dd, *J* = 5.1, 3.3 Hz, H-3), 4.08 (1H, dd, *J* = 12.0, 3.8 Hz, H-5a), 3.8 (1H, dd, *J* = 12.0, 6.8 Hz, H-5b), 3.63 (1H, d, *J* = 6.4 Hz, 1-OH), 2.11 (3H, s, OAc), 2.09 (3H, s, OAc), 2.07 (3H, s, OAc). ^{13}C NMR ($CDCl_3$, 150 MHz) δ 170.5 ($COCH_3$), 170.2 ($COCH_3$), 170.1 ($COCH_3$), 93.1 (C-1), 69.8 (C-3), 66.9, 66.8, 61.6 (C-5), 20.98 ($COCH_3$), 20.95 ($COCH_3$), 20.9 ($COCH_3$).

Minor product: ^{1}H NMR (600 MHz, $CDCl_3$) δ 5.52-5.50 (1H, H-2, obscured by major product H-2), 5.06-5.03 (2H, m, H-1 and H-4), (H-3 obs.), 4.03 (1H, dd, *J* = 11.9, 8.2 Hz, H-5a), 3.91 (1H, d, *J* = 8.9 Hz, 1-OH), 3.69 (1H, dd, *J* = 10.3, 4.1 Hz, H-5b), 2.14 (3H, s, OAc), 2.12 (3H, s, OAc), 2.05 (3H, s, OAc). ^{13}C NMR ($CDCl_3$, 150 MHz) δ 170.0($COCH_3$), 169.9($COCH_3$), 169.7($COCH_3$), 91.3 (C-1), 68.8, 67.6, 65.9, 60.6, 20.93 ($COCH_3$).

7.4.2 Glycosylations

General procedure 4.2: Reactants were dissolved in water and the pH of the solution recorded and adjusted if necessary. The solution was spread in a thin layer on a glass petri dish, and placed on a hot plate set to 100 °C. Reaction times are recorded from when evaporation was complete. The residue was periodically resuspended in D_2O and subjected to NMR analysis.

General procedure 4.3: Reactants were dissolved in water and the pH of the solution recorded and adjusted if necessary. The solution was placed in a 5 mL round bottomed flask and dried on a rotary evaporator, leaving a thin layer of residue around the inside of the flask. The flask was placed in an oil bath and heated at 100 °C. The residue was periodically resuspended in D_2O and subjected to NMR analysis.

8.4.2.1 Decomposition of reactants under glycosylation conditions

Entry	*Ribo*-30	$MgSO_4$	$MgCl_2$	solvent	pH
1[a]	0.036 mmol	0.020 mmol	0.010	H_2O, 400 µL	5.9
2[b]	0.062 mmol	0.042 mmol	0.022	100 mM Pi, 800 µL	6.0

Table 7.4.3: *Decomposition of ribose ribo-**30** under glycosylation conditions in the presence of Mg^{2+}. ^1H NMR spectra shown in **Figure 7.4.1***

[a] *General procedure 4.2.*

[b] *General procedure 4.3.*

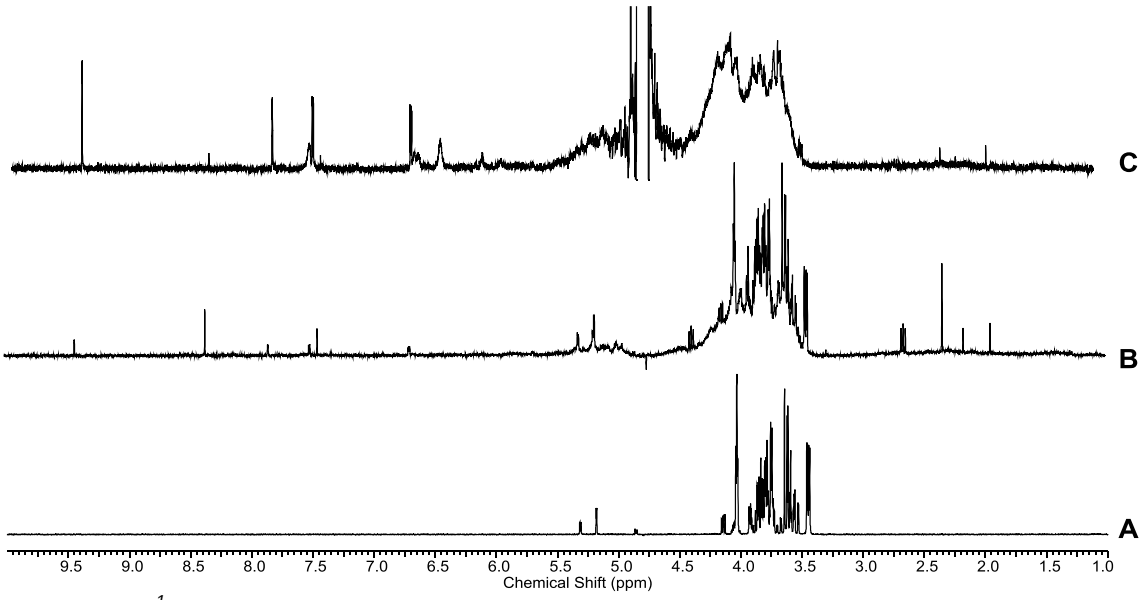


Figure 7.4.1 *^1H NMR spectra (600 MHz, D_2O, 1.0-10.0 ppm) showing decomposition of ribose ribo-**30** under glycosylation conditions. **A**) ribo-**30**; and **B**) after heating with $MgSO_4$ (1.8 eq), $MgCl_2$ (4 eq) at 100 °C for 1 h (**Table 8.4.X**, entry 1); and **C**) after heating with $MgSO_4$ (0.7 eq), $MgCl_2$ (0.35 eq) and Pi (1.3 eq.) at 100 °C for 4 h (pH 5.4, **Table 8.4.X**, entry 2).*

Entry	*Ribo*-30	Iron additive	water	Initial pH	pH after 2 h
1[a]	0.041 mmol	$FeCl_2$, 0.040 mmol	500 µL	6.3	6.1
2[b]	0.041 mmol	$FeCl_3$, 0.040 mmol	500 µL	6.3	7.1

Table 7.4.4: *Decomposition of ribose ribo-**30** under glycosylation conditions in the presence of iron, general procedure 4.2. ^{1}H NMR spectra shown in **Figure 7.4.2***

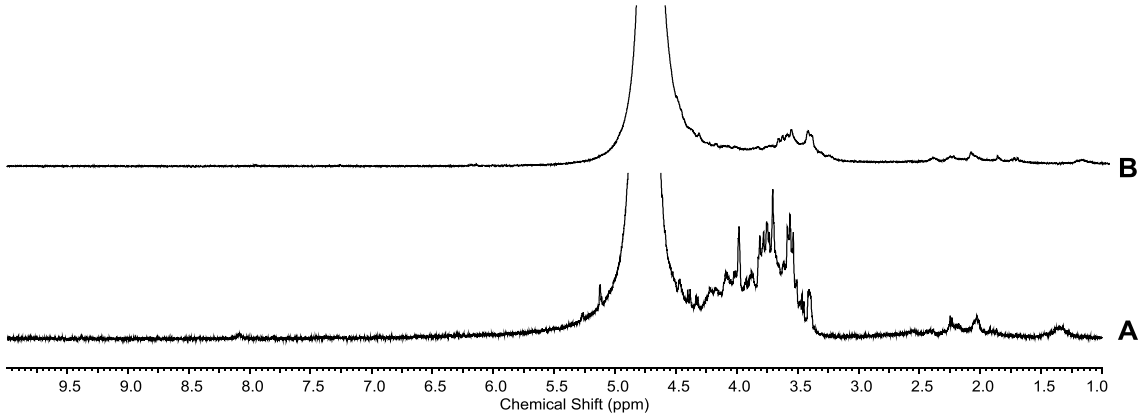


Figure 7.4.2 *^{1}H NMR spectra (600 MHz, D_2O, 1.0-10.0 ppm) showing decomposition of ribose ribo-**30** under glycosylation conditions in the presence of iron. Ribo-**30** (1 eq.) after heating for 2 h in the presence of: **A**) $FeCl_2$ (1 eq.); and **B**) $FeCl_3$ (1 eq.).*

1,2,3-Tris-acetyl-ribose-5-phosphate 170: 1,2,3-Tris-acetyl-ribose-5-phosphate **170** (0.026 mmol), $MgSO_4$ (0.022 mmol) and $MgCl_2$ (0.01 mmol) dissolved in water (2 mL, pH 5.8), general procedure 4.2. ***Figure 7.4.3.***

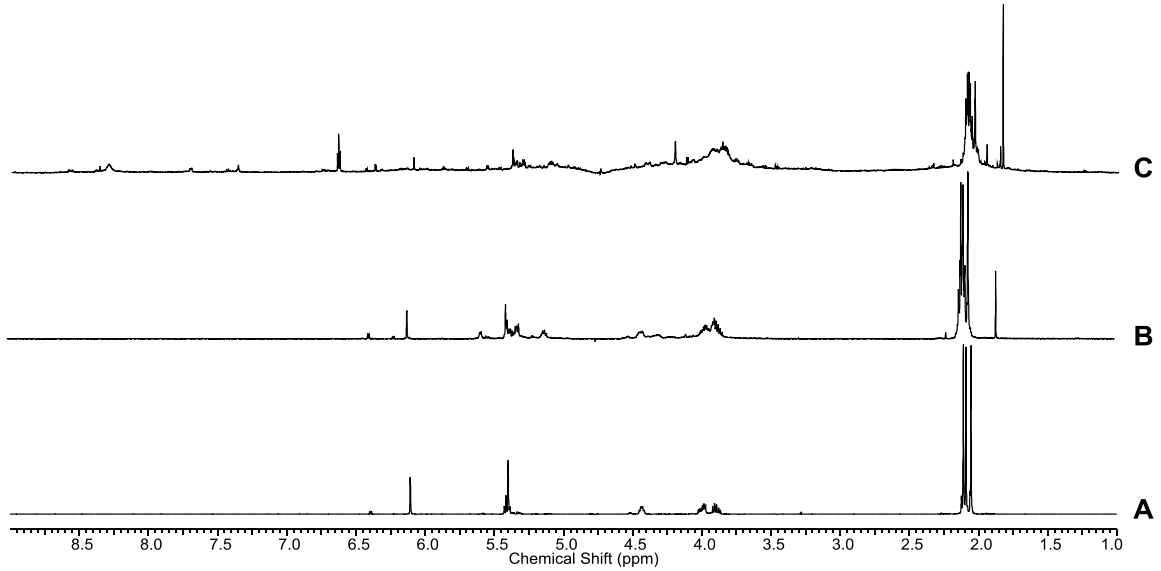


Figure 7.4.3 *^{1}H NMR spectra (600 MHz, D_2O, 1.0-9.0 ppm) showing decomposition of 1,2,3-tris-acetyl-ribose-5-phosphate **170** under glycosylation conditions. **A**) **170**; and after heating with $MgSO_4$ (1 eq) and $MgCl_2$ (0.5 eq) at 100 °C for: **B**) 2h; and **C**) 6 h.*

2,3,5-Tris-acetyl ribose 171: 2,3,5-Tris-acetyl ribose **171** (0.09 mmol), $MgSO_4$ (0.022 mmol) and $MgCl_2$ (0.01 mmol) were dissolved in water (400 µL). General procedure 4.3, but the residue was periodically resuspended in $CDCl_3$ (instead of D_2O). ***Figure 7.4.4.***

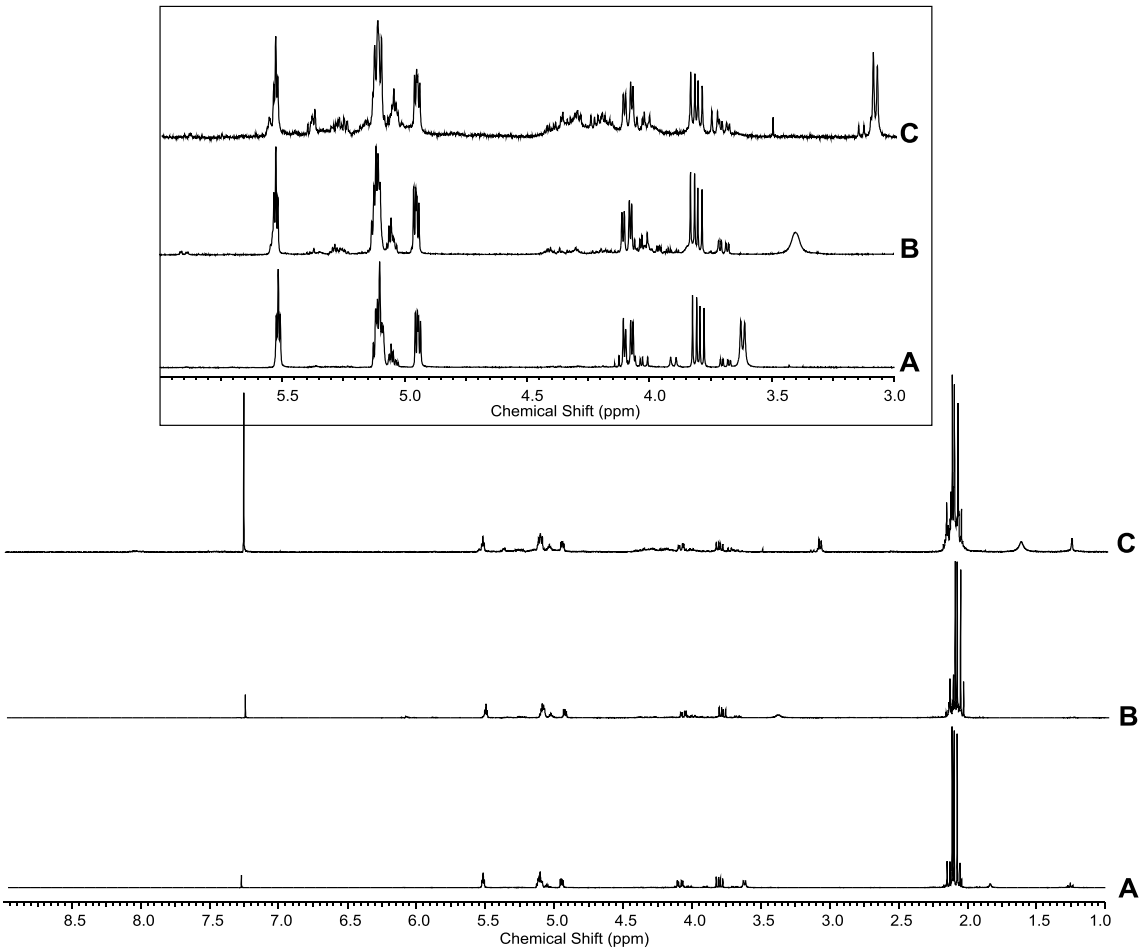


Figure 7.4.4 *^{1}H NMR spectra (600 MHz, D_2O, 1.0-9.0 ppm) showing decomposition of 2,3,5-tris-acetyl ribose **37** under glycosylation conditions. **A**) 2,3,5-Tris-acetyl ribose **171** (9 eq.); and after heating with $MgSO_4$ (2 eq.) and $MgCl_2$ (1 eq.) at 100 °C for: **B**) 2 h; and **C**) 41 h.*

2,3,5-Tris-acetyl ribose **171** (0.0066 mmol), $MgSO_4$ (0.041 mmol) and $MgCl_2$ (0.02 mmol) were dissolved in water (800 µL, pH 6.3), general procedure 4.3. ***Figure 7.4.5.***

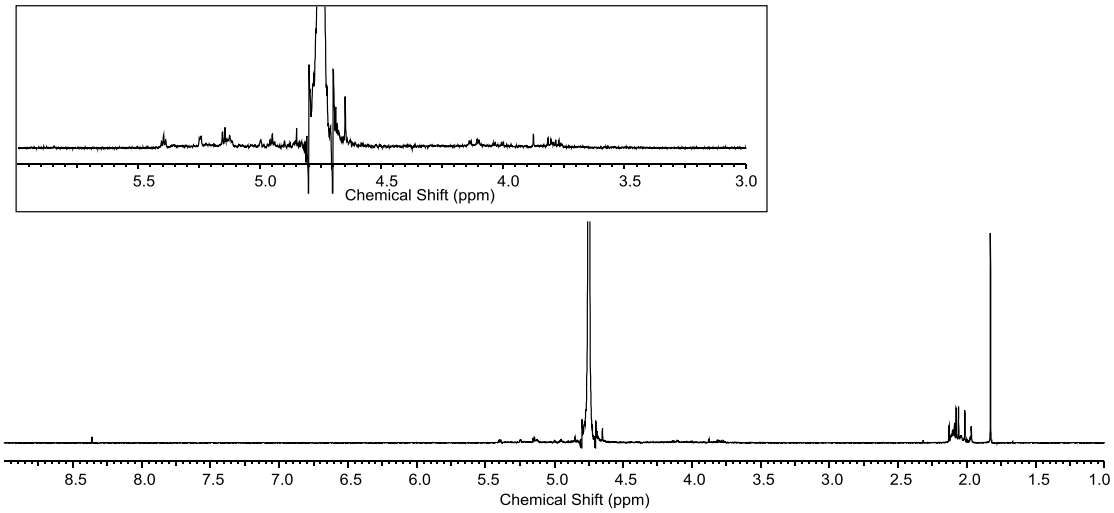


Figure 7.4.5 *^{1}H NMR spectrum (600 MHz, D_2O, 1.0-9.0 ppm; inset 3.0-6.0 ppm) showing decomposition of 2,3,5-tris-acetyl ribose* ***171*** *under glycosylation conditions.* ***171*** *(3.3 eq.) after heating with $MgSO_4$ (2.1 eq.) and $MgCl_2$ (1 eq.) at 100 °C for 96 h (pH 10.2).*

8.4.2.2 Attempted glycosylation reactions using purine nucleobases

Hypoxanthine **153** (0.002 mmol, 1 eq.), $MgSO_4$ (0.021 mmol, 10.5 eq.), $MgCl_2$ (0.01 mmol, 5 eq.) and *ribo*-**30** (0.04 mmol, 20 eq.) were dissolved in water (400 µL, pH 6.2), general procedure 4.3. Additional hypoxanthine **153** (0.002 mmol, 1 eq.) was added after 2 h. After 4 h heating, the NMR sample was spiked with inosine **57**, confirming the presence of this nucleoside (***Figure 4.14***).

1,2,3-Tris-acetyl-ribose-5-phosphate 170 and hypoxanthine 153

Entry	153	170	$MgSO_4$	$MgCl_2$	H_2O	pH	Heat time
1[a]	0.005 mmol, 1 eq.	0.07 mmol, 14 eq.	0.02 mmol, 4 eq.	0.04 mmol, 8 eq.	800 µL	6.0	4 h
2[b]	0.002 mmol, 1 eq.	0.03 mmol, 15 eq.	0.021 mmol, 10.5 eq.	0.01 mmol, 5 eq.	400 µL	6.0	10 h

Table 7.4.5: *Attemped glycosylation of hypoxanthine* ***153*** *with 1,2,3-tris-acetyl-ribose-5-phosphate* ***170*** *under glycosylation conditions in the presence of Mg^{2+}, general procedure 4.3. No nucleoside observed by ^{1}H NMR analysis; ^{1}H NMR spectra for entry 2 shown in* ***Figure 7.4.6.***

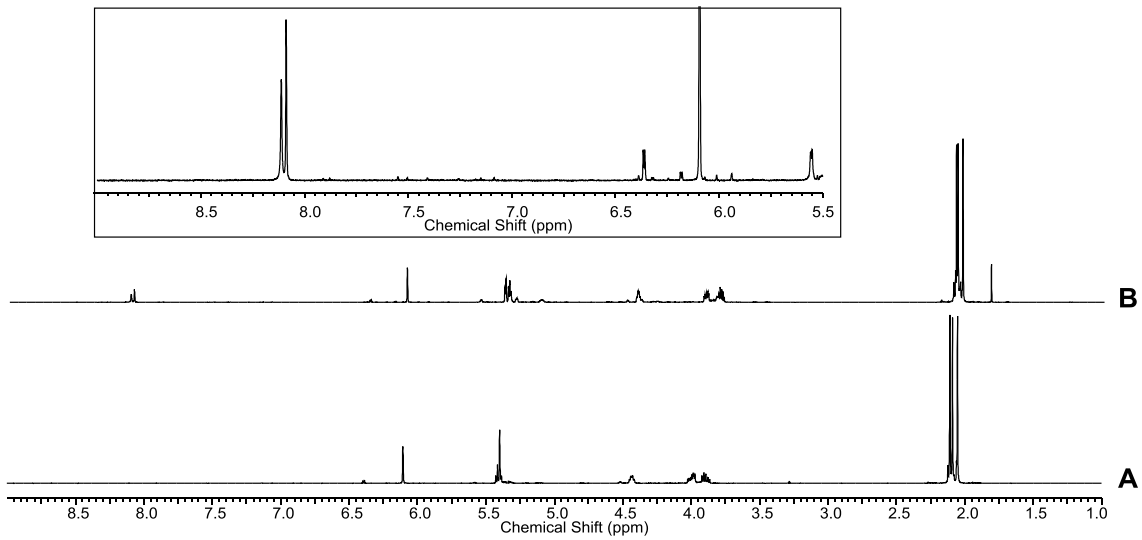


Figure 7.4.6 *^{1}H NMR spectra (600 MHz, D_2O, 1.0-9.0 ppm; inset 5.5-9.0 ppm) showing attempted glycosylation of hypoxanthine **153** with 1,2,3-tris-acetyl-ribose-5-phosphate **170**. **A**) **170** (15 eq.); and **B**) after heating at 100 °C with **153** (1 eq.), $MgSO_4$ (10.5 eq.), $MgCl_2$ (5 eq.) for 10 h. A number of small peaks are visible in the H-1' region (δ 6.0-6.5) there is no corresponding aromatic peak. The two – apparently unequal – peaks in the aromatic region in fact integrate 1:1 and are **153**.*

2,3,5-Tris-acetyl-ribose 171 and hypoxanthine 153

Entry	153	$MgSO_4$	$MgCl_2$	171	H_2O	pH	Heat time
1[a, b]	1.25 µmol, 1 eq, plus 2 µmol, 1.6 eq after 2 h	10.5 µmol, 8 eq	5.0 µmol, 4 eq.	36.0 µmol, 29 eq	400 µL	5.8	4 h
2[a]	0.004 mmol, 1 eq. plus 0.004 mmol, 1 eq. after 24 h	0.042 mmol, 10.5 eq	0.02 mmol, 5 eq.	0.09 mmol, 23 eq.	900 µL	6.3	41 h
3[c, d]	0.004 mmol, 1 eq.	0.042 mmol, 10.5 eq	0.02 mmol, 5 eq	0.066 mmol, 17 eq	800 µL	6.2	96 h

Table 7.4.6: *Attempted glycosylation reactions of hypoxanthine **153** with 2,3,5-tris-acetyl-ribose **171, g**eneral procedure 4.3. Periodically analysed by NMR; no nucleoside observed.*

[a] resuspended in $CDCl_3$ for NMR analysis (2',3',5'-triacetylinosine – the desired product – is a literature compound and is soluble in $CDCl_3$)

*[b] ^{1}H NMR spectrum shown in **Figure 7.4.7***

[c] resuspended in D_2O for NMR analysis

*[d] ^{1}H NMR spectrum shown in **Figure 7.4.8***

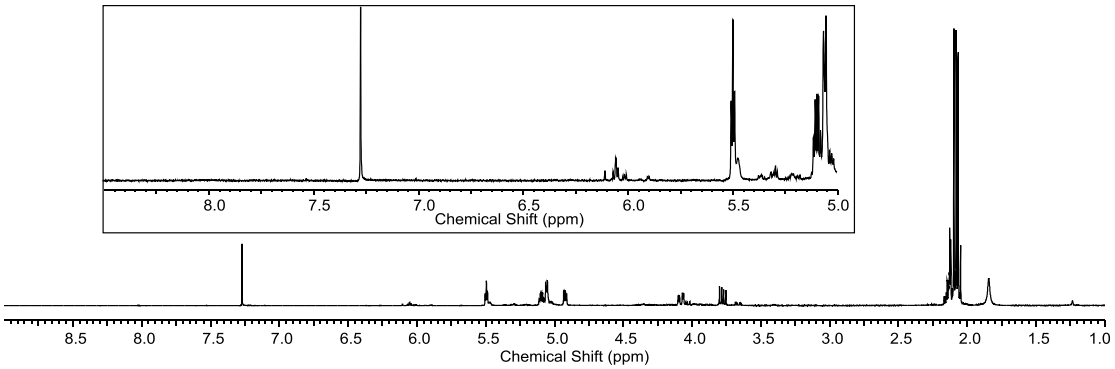


Figure 7.4.7 *^{1}H NMR spectra (600 MHz, $CDCl_3$, 1.0-9.0 ppm; inset 5.0-8.5 ppm) showing attempted glycosylation of hypoxanthine **153** with 2,3,5-tris-acetyl-ribose **171** (**Table 7.4.6**, entry 1). A number of small peaks can be seen in the H-1' region (δ 6.0-6.5), but there is no corresponding aromatic peak. (**153** is not soluble in $CDCl_3$).*

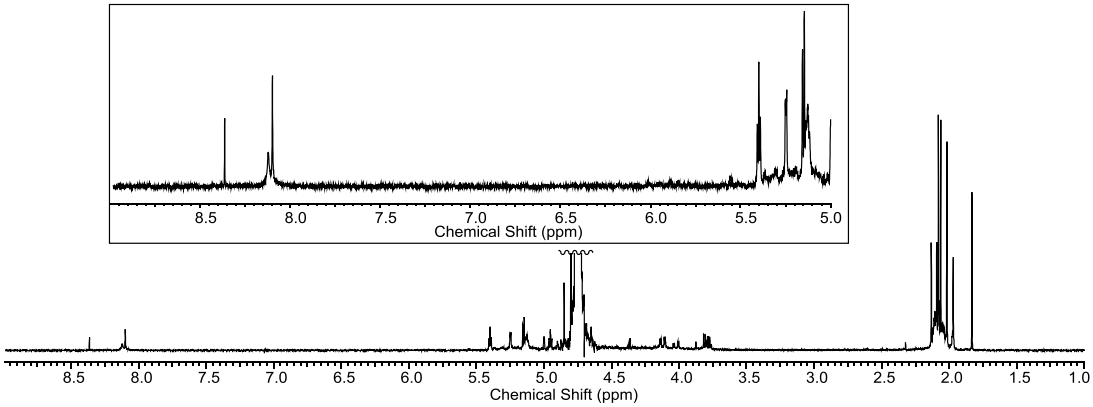


Figure 7.4.8 *^{1}H NMR spectrum (600 MHz, D_2O, 1.0-9.0 ppm; inset 5.0-9.0 ppm) showing attempted glycosylation of hypoxanthine **153** with 2,3,5-tris-acetyl-ribose **171** (**Table 7.4.6**, entry 3). The peak at δ 8.36 is a decomposition product of the sugar and can also be seen in the control decomposition of 2,3,5-tris-acetyl-ribose **171** under glycosylation conditions (**Figure 7.4.5**).*

Entry	153	Fe additive	171	Initial pH	Final pH
1	0.0025 mmol, 1 eq.	$FeCl_2$ (0.04 mmol, 16 eq,)	0.04 mmol, 16 eq.	6.3	3.4
2	0.0025 mmol, 1 eq.	$FeCl_3$ (0.04 mmol, 16 eq.)	0.04 mmol, 16 eq.	6.0	4.4

Table 7.4.7: *Attempted glycosylation reactions of hypoxanthine **153** with 2,3,5-tris-acetyl-ribose **171** in the presence of iron additives**,** general procedure 4.3. Heated for 12 h and periodically analysed by ^{1}H NMR; no nucleoside observed.*

2-Deoxyribose 176 and adenine 44: Adenine **44** (0.005 mmol, 1 eq.), $MgSO_4$ (0.043 mmol, 8.5 eq.), $MgCl_2$ (0.022 mmol, 4.5 eq.) and 2-deoxyribose **176** (0.064 mmol, 13 eq.) were dissolved in water (800 µL) and the solution adjusted from pH 6.5 to 6.1, general procedure 4.3. After 7 h heating (pH 8.2) a number of new peaks in the aromatic region were seen, as well as a new peak in the H-1′ region, but negative spike with 2′-deoxyadenosine **174**.

Entry	base	*ribo*-**30**-5P	$MgSO_4$	$MgCl_2$	pH	Heat time
1[a]	**153** (0.005 mmol, 1 eq.)	0.062 mmol, 12 eq.	0.045 mmol, 9 eq.	0.023 mmol, 4.6 eq.	6.0	17 h
2	**153** (0.005 mmol, 1 eq.)	0.062 mmol, 12 eq.	0.045 mmol, 9 eq.	0.023 mmol, 4.6 eq.	6.0	2 h (pH 6.7)
3[b]	**44** (0.005 mmol, 1 eq.)	0.062 mmol, 12 eq.	0.043 mmol, 8.5 eq.	0.022 mmol, 4.5 eq.	6.2	7 h (pH 6.6)

Table 7.4.8: *Attempted glycosylation reactions of hypoxanthine **153** and adenine **44** with ribose-5-phosphate ribo-**30**-5P, general procedure 4.3, 800 µL H_2O.*

[a] *negative spike with inosine-5-phosphate ß-**57**-5'P*

[b] *negative spike with adenosine-5′-phosphate ß-**55**-5'P*

Arabinose *ara*-**30 and adenine 44:** Adenine **44** (0.008 mmol, 1 eq.), $MgSO_4$ (0.063 mmol, 7.9 eq.), $MgCl_2$ (0.034 mmol, 4.3 eq.) and arabinose *ara*-**30** (0.101 mmol, 12.6 eq.) in water (350 µL), general procedure 4.3. After 4 h heating a number of new peaks were observed in the aromatic region and in the H-1′ region, but negative spike with arabinose adenosine *ara*-**55**.

Attempted glycosylation reactions of hypoxanthine 153 in the presence of inorganic phosphate

Entry	153	Ribose derivative	$MgSO_4$	$MgCl_2$	pH	Heat time
1[a,]	0.005 mmol, 1 eq.	*ribo*-**30** (0.063 mmol, 13 eq.)	0.044 mmol, 8.4 eq.	0.022 mmol, 4.4 eq.	6.2	4 h (pH 6.2)
2	0.005 mmol, 1 eq.	*ribo*-**30**-5P (0.058 mmol, 12 eq.)	0.044 mmol, 8.4 eq..	0.022 mmol, 4.4 eq.	6.2	8 h (pH 6.1)
3	0.005 mmol, 1 eq.	**176** (0.068 mmol, 14 eq.)	0.044 mmol, 8.4 eq.	0.022 mmol, 4.4 eq.	6.2	8 h (pH 5.6)

Table 7.4.9: *Attempted glycosylation reactions of hypoxanthine **153** with ribose ribo-30, ribose-5-phosphate ribo-**30**-5P and 2-deoxyribose **176**, general procedure 4.3, in the presence of 100 mM Pi (800 µL, 16 eq.). No nucleoside observed by ^{1}H NMR analysis.*

[a] *negative spike with inosine ß-**57** (**Figure 7.4.9**).*

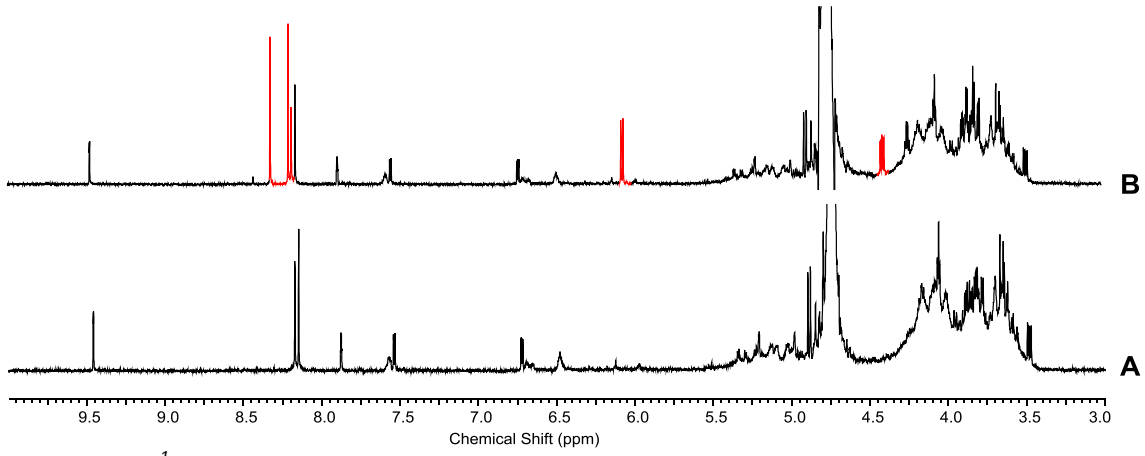


Figure 7.4.9 *^{1}H NMR spectra (600 MHz, D_2O, 3.0-10.0 ppm) showing attempted glycosylation of hypoxanthine **153** with ribo-**30** in the presence of inorganic phosphate. **A**) **153** (1 eq.), $MgSO_4$ (8.4 eq.), $MgCl_2$ (4.4 eq.) and ribo-**30** (13 eq.) after heating at 100 °C for 4 h in the presence of inorganic phosphate (16 eq.); and **B**) negative spike with inosine **57**.*

8.4.3: Decomposition of purine nucleosides and nucleotides under glycosylation conditions

Inosine ß-57

Entry	ß-57	$MgSO_4$	$MgCl_2$	Solvent	pH	Heat time
1[a]	0.069 mmol	0.041 mmol	0.020 mmol	H_2O, 800 µL	6.1	26 h (pH 3.8)
2[b]	0.068 mmol	0.042 mmol	0.022 mmol	100 mM Pi, 800 µL	6.0	5 h (pH 5.9)

Table 7.4.10: *Decomposition of inosine ß-**57** under glycosylation conditions, general procedure 4.3. New aromatic peaks observed by ^{1}H NMR analysis in both cases*

*[a] negative spike with hypoxanthine **153** (**Figure 7.4.10**).*

*[b] positive spike with hypoxanthine **153** (new singlet peaks at δ 8.09 and 8.06; 19% conversion by relative integration).*

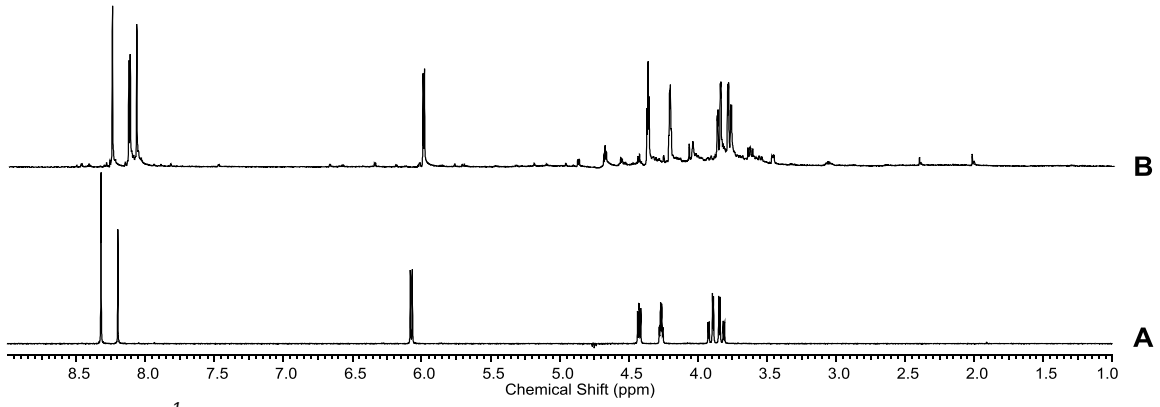


Figure 7.4.10 *^{1}H NMR spectra (600 MHz, D_2O, 1.0-9.0 ppm) showing decomposition of inosine **57** under glycosylation conditions. **A**) **57** (3.5 eq.); and **B**) after heating with $MgSO_4$ (2 eq.) and $MgCl_2$ (1 eq.) at 100 °C for 26 h.*

Entry	ß-57	Iron additive	H_2O	pH	Heat time
1	0.044 mmol	$FeCl_2$ (0.040 mmol)	500 µL	6.2	2 h (pH 6.1)
2	0.044 mmol	$FeCl_3$ (0.040 mmol)	500 µL	6.2	2 h (pH 6.8)

Table 7.4.11: *Decomposition of inosine ß-**57** in the presence of iron additives, general procedure 4.3. Only one species (inosine ß-**57**) observed ^{1}H NMR analysis in each case.*

2′,3′,5′-Tri-acetylated inosine 177[b]

Entry	177	$MgSO_4$	$MgCl_2$	Solvent	pH	Heat time
1[a]	0.03 mmol, 3 eq.	0.02 mmol, 2 eq.	0.01 mmol, 1 eq.	H_2O, 400 µL	6.0	30 h (pH 7.5)
2[a]	0.04 mmol, 4 eq.	0.02 mmol, 2 eq.	0.01 mmol, 1 eq.	100 mM Pi, 400 µL, 4 eq.	6.0	30 h (pH 6.3)
3[b]	0.03 mmol, 3 eq.	0.02 mmol, 2 eq.	0.01 mmol, 1 eq.	175 mM Pi, 400 µL, 6.4 eq.	6.1	12 h (pH 6.3)
4[a]	0.03 mmol, 3 eq.	0.02 mmol, 2 eq.	0.01 mmol, 1 eq.	175 mM Pi, 400 µL, 6.4 eq.	6.3	16 h (pH 6.4)
5[a]	0.03 mmol, 3 eq.	0.02 mmol, 2 eq.	0.01 mmol, 1 eq.	175 mM Pi, 400 µL, 6.4 eq.	6.3	16 h (pH 6.4)

Table 7.4.12: *Decomposition* of *2′,3′,5′-tri-acetylated inosine **177** under glycosylation conditions, general procedure 4.3. New aromatic peaks observed by ^{1}H NMR analysis in both cases*

[a] *only one species (**177**) observed by ^{1}H NMR analysis.*

[b] *After heating for 8 h (pH 6.4) only one species (**177**) observed by ^{1}H NMR analysis. After 12 h total heating (pH 5.9), ^{1}H NMR analysis revealed two nucleoside species present in a 1:1 ratio, but only five acetyl groups.*

Entry	177	Iron additive	H_2O	pH	Heat time
1	0.04 mmol, 1 eq.	$FeCl_2$ (0.04 mmol, 1 eq.)	500 µL	6.2	12 h (pH 7.1)
2	0.04 mmol, 1 eq.	$FeCl_3$ (0.04 mmol, 1. eq)	500 µL	6.1	12 h (pH 8.4)

Table 7.4.13: *Decomposition of 2′,3′,5′-tri-acetylated inosine **177** in the presence of iron additives, general procedure 4.3. Only one species (**177**) observed ^{1}H NMR analysis in each case.*

2′,3′-Di-acetylated inosine-5′-monophosphate 178[c]

2′,3′-Di-acetylated inosine-5′-monophosphate **178** (0.03 mmol, 3 eq.), $MgSO_4$ (0.02 mmol, 2 eq.) and $MgCl_2$ (0.01 mmol, 1 eq.) were dissolved in water (400 µL, pH 6.1), general procedure 4.3. No change was observed by ^{1}H NMR analysis after heating for 30 h (pH 6.3).

[b] 2′,3′,5′-triacetylated inosine **45** prepared by Dr C.A. Fernandez-Garcia

[c] 2′,3′-Di-acetylated inosine-5′-monophosphate **178** prepared by Dr C.A. Fernandez-Garcia

2′,3′-Di-acetylated inosine-5′-monophosphate **178** (0.035 mmol, 3.2 eq.), $MgSO_4$ (0.02 mmol, 2 eq.), $MgCl_2$ (0.01 mmol, 1 eq.) were dissolved in 25 mM phosphate buffer (400 µL, 0.9 eq., pH 5.7), general procedure 4.3. **178** decomposed to give a variety of other products over time (***Table 7.4.14)***; conversion estimated based on relative integration of H-1′ peaks.

Time / h	Yield by relative ^{1}H NMR integration							pH
	δ 6.35 (**178**)	δ 6.31	δ 6.27	δ 6.23	δ 6.14 (ß-**57**-5'P)	δ 6.09	δ 6.07	
2	92%	3%	2%	-	3%	-	-	6.7
4	82%	6%	4%	-	8%	-	-	6.6
8	69%	9%	3%	-	12%	1%	1%	6.5
12	49%	13%	10%	2%	19%	4%	3%	5.7

Table 7.4.14: *Yields of various products (identified by H-1′ chemical shift) formed from the decomposition of 2',3'-di-acetylated inosine-5'-monophosphate* ***178*** *under glycosylation conditions in the presence of Pi (0.9 eq).*

2′,3′-Di-acetylated inosine-5′-monophosphate **178** (0.03 mmol, 3 eq.), $MgSO_4$ (0.02 mmol, 2 eq.), $MgCl_2$ (0.01 mmol, 1 eq.) were dissolved in 100 mM phosphate buffer (400 µL, 4 eq., pH 6.0), general procedure 4.3. **178** decomposed to give a variety of other products (***Figure 7.4.11, Table 7.4.15***); conversion estimated based on relative integration of H-1′ peaks.

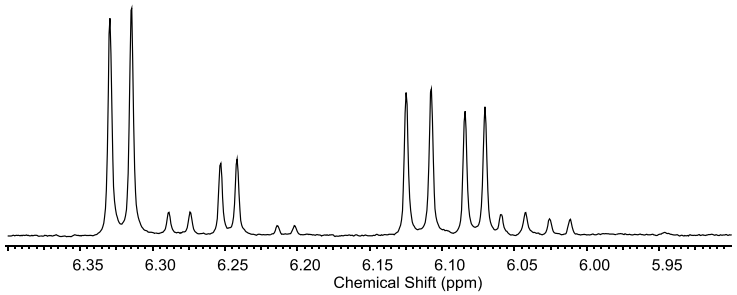


Figure 7.4.11 *^{1}H NMR spectrum (600 MHz, D_2O, 5.80-6.40 ppm) showing 8 species observed containing an anomeric proton after heating 2',3'-di-acetylated inosine-5'-monophosphate* ***178*** *(3 eq.), $MgSO_4$ (2 eq.) and $MgCl_2$ (1 eq.) at 100 °C for 2 h in the presence of inorganic phosphate (4 eq.). δ 6.35 (36%, d, J = 6.4 Hz, spike positive for* ***178****), 6.31 (3%, d, J = 5.4 Hz), 6.27 (12%, d, J = 4.6 Hz), 6.23 (1%, d, J = 5.2 Hz), 6.14 (23%, d, J = 6.6 Hz, spike positive for inosine-5'-monophosphate ß-***57****-5'P), 6.09 (20%, d, J = 6.0 Hz), 6.07 (3%, d, J = 6.7 Hz), 6.04 (2%, d, J = 6.0 Hz, spike positive for inosine* ***153****)*

Time / h	Yield by relative ^{1}H NMR integration								pH
	δ 6.35 (**178**)	δ 6.31	δ 6.27	δ 6.23	δ 6.14 (ß-**57**-5'P)	δ 6.09	δ 6.07	δ 6.04 (**57**)	
2	36%	3%	12%	1%	23%	20%	3%	2%	5.5

Table 7.4.15: *Yields of various products (identified by H-1′ chemical shift) formed from the decomposition of 2',3'-di-acetylated inosine-5'-monophosphate* ***178*** *under glycosylation conditions in the presence of Pi (4 eq).*

The decomposition of 2′,3′-di-acetylated inosine-5′-monophosphate **178** under glycosylation conditions with a sub-stoichiometric quantity of inorganic phosphate was repeated with variations in the initial pH value:

2′,3′-Di-acetylated inosine-5′-monophosphate **178** (0.033 mmol, 3.2 eq.), $MgSO_4$ (0.02 mmol, 2 eq.) and $MgCl_2$ (0.01 mmol, 1 eq.) were dissolved in 25 mM phosphate buffer (400 µL, 0.9 eq., pH 6.0), general procedure 4.3. 2′,3′-Di-acetylated inosine-5′-monophosphate **178** decomposed to give a variety of other products (***Table 7.4.16***); conversion estimated based on relative integration of H-1′ peaks.

Time / h	Yield by relative ^{1}H NMR integration							pH
	δ 6.35 (**178**)	δ 6.31	δ 6.27	δ 6.23	δ 6.14 (ß-**57**-5'P)	δ 6.09	δ 6.07	
12	70%	13%	4%	1%	10%	1%	1%	5.9
14	69%	14%	4%	1%	10%	1%	1%	5.9
16	68%	14%	5%	1%	10%	1%	1%	5.9

Table 7.4.16: *Yields of various products (identified by H-1′ chemical shift) formed from the decomposition of 2′,3′-di-acetylated inosine-5′-monophosphate **178** under glycosylation conditions in the presence of Pi (0.9 eq).*

2′,3′-Di-acetylated inosine-5′-monophosphate **178** (0.033 mmol, 3.2 eq.), $MgSO_4$ (0.02 mmol, 2 eq.) and $MgCl_2$ (0.01 mmol, 1 eq.) were dissolved in 25 mM phosphate buffer (400 µL, 0.9 eq., pH 6.3), general procedure 4.3. **178** decomposed to give a variety of other products (***Table 7.4.17***); conversion was estimated based on relative integration of H-1′ peaks.

Time / h	^{1}H NMR integration relative to other species integrated by H-1′ peak							pH
	δ 6.35 (**178**)	δ 6.31	δ 6.27	δ 6.23	δ 6.14 (ß-**57**-5'P)	δ 6.09	δ 6.07	
12	89%	5%	2%	-	4%	-	-	6.6
14	87%	6%	2%	-	5%	-	-	6.5
16	85%	7%	3%	-	5%	<1%	-	6.5

Table 7.4.17: *Yields of various products (identified by H-1′ chemical shift) formed from the decomposition of 2′,3′-di-acetylated inosine-5′-monophosphate **178** under glycosylation conditions in the presence of Pi (0.9 eq).*

		Yield by relative ^{1}H NMR integration								
Pi / eq	Initial pH	δ 6.35 (**178**)	δ 6.31	δ 6.27	δ 6.23	δ 6.14 (ß-**57**-5'P)	δ 6.09	δ 6.07	δ 6.04 (ß-**57**)	Final pH
0.37	6.3	18%	14%	8%	6%	16%	14%	14%	9%	5.4
1.3	6.0	23%	20%	7%	6%	14%	11%	13%	8%	5.9

Table 7.4.18: *Yields of various products (identified by H-1' chemical shift) formed from the decomposition of 2',3'-di-acetylated inosine-5'-monophosphate* ***178*** *(0.027 mmol) under glycosylation conditions in the presence of Pi (400 µL) after heating for 2 h, general procedure 4.3.* ^{1}*H NMR spectra shown in* ***Figure 7.4.12****.*

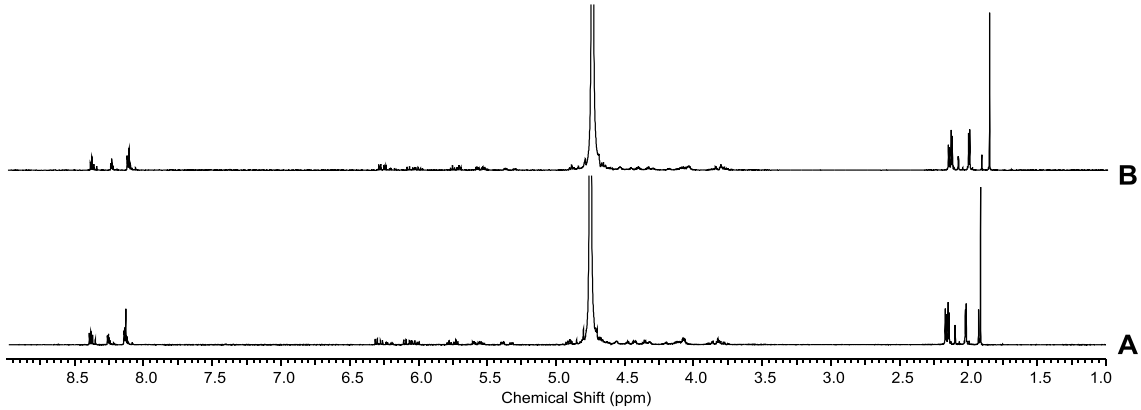


Figure 7.4.12 ^{1}*H NMR spectra (600 MHz,* D_2O*, 1.0-9.0 ppm) showing decomposition of 2',3'-di-acetylated inosine-5'-monophosphate* ***178*** *in the presence of phosphate after 2 h at 100 °C.* ***A****) with 0.37 eq Pi; and* ***B****) with 1.3 eq. Pi.*

Inosine-5-phosphate ß-**57**-5'P

Entry	ß-**57**-5'P	$MgSO_4$	$MgCl_2$	Solvent	pH	Heat time
1[a]	0.071 mmol	0.041 mmol	0.020 mmol	H_2O, 800 µL	6.3	50 h (pH 6.7)
2[b]	0.070 mmol	0.042 mmol	0.022 mmol	100 mM Pi, 800 µL	6.0	5 h (pH 6.3)

Table 7.4.19: *Decomposition* of *inosine-5-phosphate ß-**57**-5'P under glycosylation conditions, general procedure 4.3.*

[a] *Positive spike with inosine* ***153*** *(5% estimated conversion by relative integration of H-1' peaks) (****Figure 7.4.13****).*

[b] *Three species observed by* ^{1}*H NMR (****Figure 7.4.14****); yields estimated by relative integration of aromatic peaks: ß-**57**-5'P (67%), ß-**57** (19%, positive spike), and* ***153*** *(14%).*

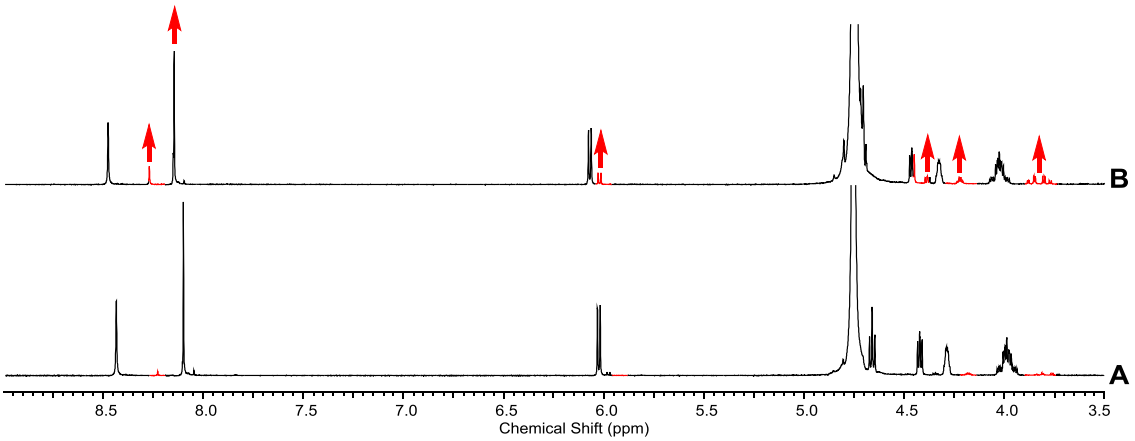


Figure 7.4.13: *^{1}H NMR spectra (600 MHz, D_2O, 1.0-9.0 ppm) showing decomposition of inosine-5-phosphate ß-**57**-5'P under glycosylation conditions.* ***A****) ß-**57**-5'P (3.5 eq.) after heating with $MgSO_4$ (2 eq.) and $MgCl_2$ (1 eq.) at 100 °C for 50 h; and* ***B****) spiked with inosine **57**.*

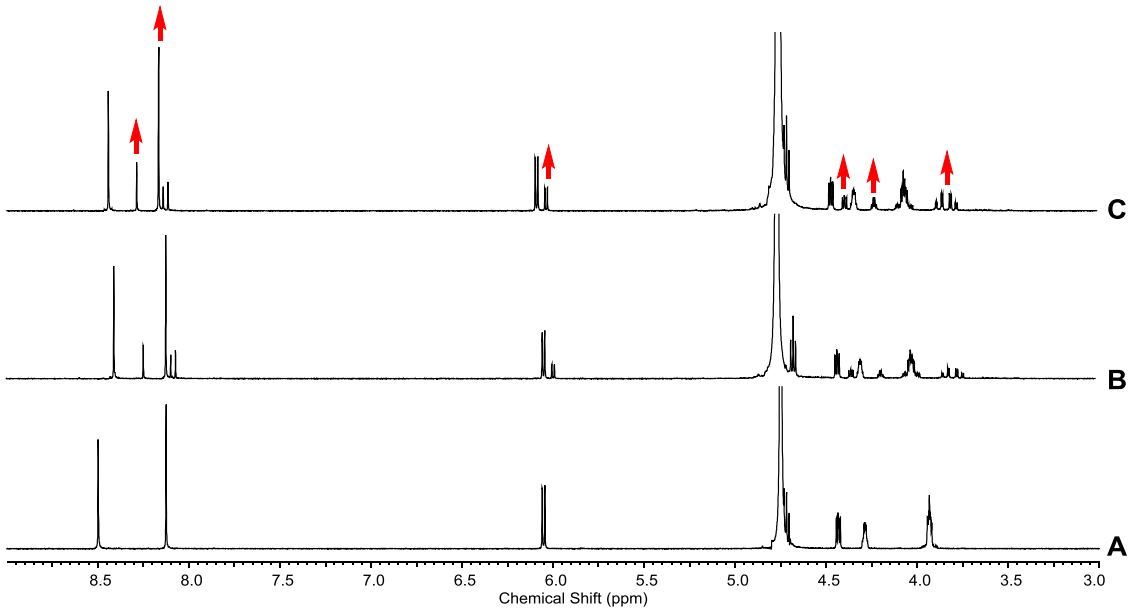


Figure 7.4.14 *^{1}H NMR spectra (600 MHz, D_2O, 3.0-9.0 ppm) showing decomposition of inosine-5-phosphate ß-**57**-5'P under glycosylation conditions in the presence of phosphate buffer.* ***A****) ß-**57**-5'P (3.2 eq.);* ***B****) after heating with $MgSO_4$ (1.6 eq.), $MgCl_2$ (1 eq.) and Pi (3.6 eq.) at 100 °C for 50 h; and* ***C****) spiked with inosine ß-**57**.*

Adenosine ß-**55** (0.059 mmol), $MgSO_4$ (0.041 mmol) and $MgCl_2$ (0.02 mmol) were dissolved in water (800 µL) and adjusted from pH 6.8 to 6.1, general procedure 4.3. Only ß-**55** could be observed by ^{1}H NMR analysis (***Figure 7.4.15***) after heating for 24 h (pH 9.7).

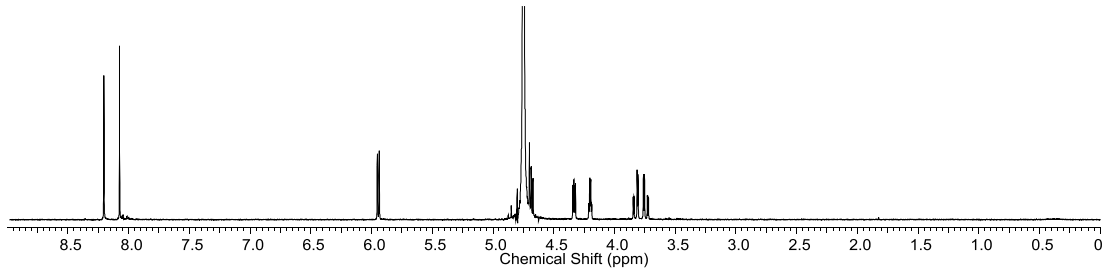


Figure 7.4.15 *^{1}H NMR spectrum (600 MHz, D_2O, 0.0-9.0 ppm) showing decomposition of adenosine ß-**55** under glycosylation conditions. **5** (3 eq.) after heating with $MgSO_4$ (2 eq.) and $MgCl_2$ (1 eq.) at 100 °C for 24 h.*

Adenosine-5′-monophosphate ß-**55**-5'P (0.063 mmol), $MgSO_4$ (0.041 mmol) and $MgCl_2$ (0.02 mmol) were dissolved in water (800 µL) and adjusted from pH 8.5 to 6.2, general procedure 4.3. After heating for 7 h (pH 6.2) two species could be observed by ^{1}H NMR analysis (***Figure 7.4.16***): ß-**55**-5'P and ß-**55** (6% conversion estimated by relative integration of H-1′ peaks; presence confirmed by positive spike with ß-**55**).

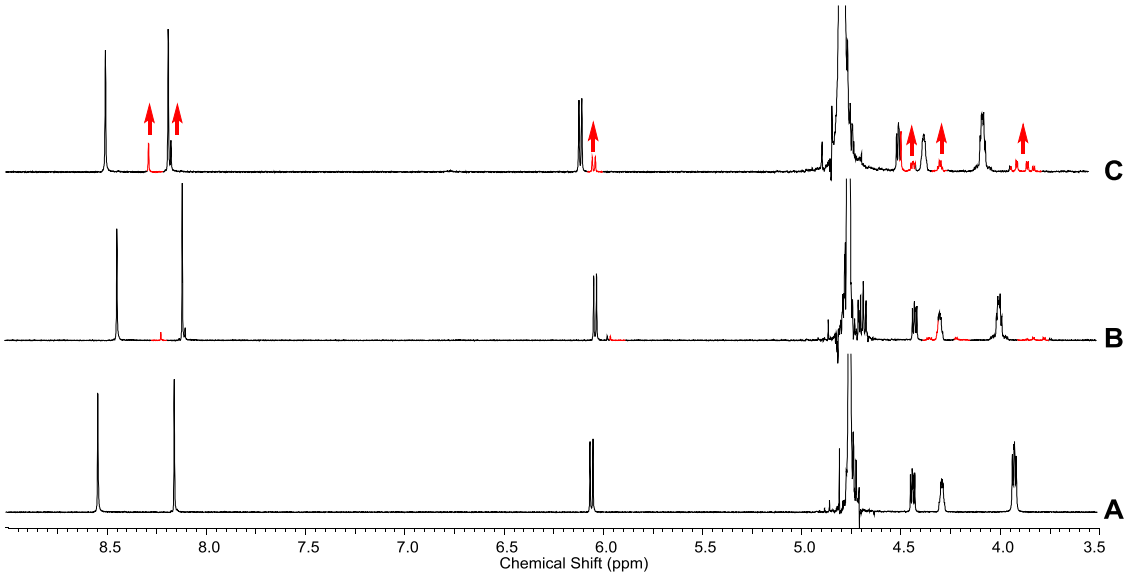


Figure 7.4.16 *^{1}H NMR spectra (600 MHz, D_2O, 3.0-9.0 ppm) showing decomposition of adenosine-5′-monophosphate ß-**55**-5'P under glycosylation conditions. **A**) ß-**55**-5'P (3.2 eq.); **B**) after heating with $MgSO_4$ (2 eq.) and $MgCl_2$ (1 eq.) at 100 °C for 24 h; and **C**) spiked with adenosine ß-**55**.*

2′-Deoxyadenosine **174** (0.060 mmol), $MgSO_4$ (0.041 mmol) and $MgCl_2$ (0.02 mmol) were dissolved in water (800 µL) and adjusted from pH 6.7 to 6.0, general procedure 4.3. After heating for 24 h (at which point the solution was pH 8.1) several new aromatic peaks are seen by ^{1}H NMR analysis (***Figure 7.4.17***).

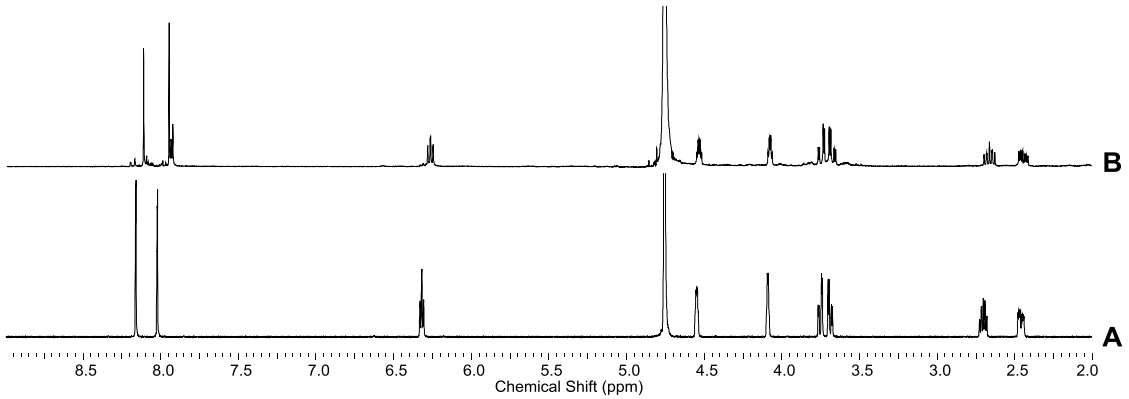


Figure 7.4.17 *^{1}H NMR spectra (600 MHz, D_2O, 3.0-9.0 ppm) showing decomposition of 2'-deoxyadenosine **174** under glycosylation conditions. **A**) **174** (3 eq.); and **B**) after heating with $MgSO_4$ (2 eq.) and $MgCl_2$ (1 eq.) at 100 °C for 24 h.*

Arabinose adenosine *ara*-**55** (0.006 mmol), $MgSO_4$ (0.063 mmol) and $MgCl_2$ (0.034 mmol) were dissolved in water (300 µL), general procedure 4.3. After heating for 4 h several new peaks were observed by ^{1}H NMR analysis (***Figure 7.4.18***).

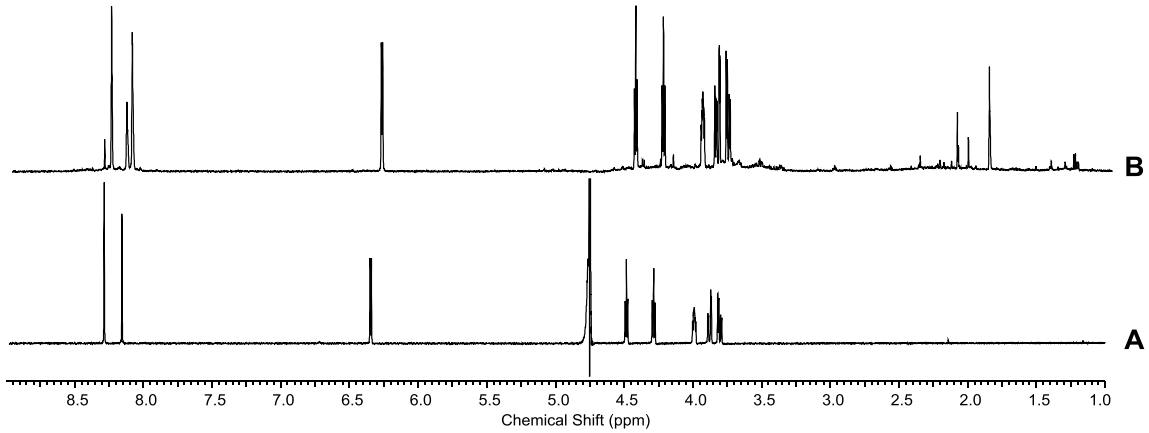


Figure 7.4.18 *^{1}H NMR spectra (600 MHz, D_2O, 3.0-9.0 ppm) showing decomposition of arabinose adenosine ara-**55** under glycosylation conditions. **A**) ara-**55** (1 eq.); and **B**) after heating with $MgSO_4$ (10.5 eq.) and $MgCl_2$ (5.7 eq.) at 100 °C for 4 h.*

8.4.4 Glycosylation of Hilbert-Johnson bases

Attempted glycosylation of 4-amino-2-methoxypyrimidine 180 with ribose *ribo*-30

Entry	Initial pH	After 2 h 100 °C			After 4 h 100 °C		
		pH	**180**	**31**	pH	**180**	**31**
1[a]	7.0	4.9	42%	58%	5.4	0%	100%
2[a]	6.5	4.6	43%	57%	5.4	25%	75%
3[a]	6.1	4.5	82%	18%	5.2	28%	72%
4[b]	5.6	5.1	44%	56%	5.3	24%	76%
5[b]	5.1	4.9	43%	57%	5.0	24%	76%
6[b]	4.6	4.8	40%	60%	5.0	19%	81%

Table 7.4.19: *Relative yields (by ^{1}H NMR integration) of 4-amino-2-methoxypyrimidine **180** and cytosine **31** observed during attempted glycosylation of **180** with ribo-**30**. No nucleoside products could be detected by ^{1}H NMR analysis.*

*[a] Ribo-**30** (0.069 mmol, 14 eq.), 4-amino-2-methoxypyrimidine **180** (0.005 mmol, 1 eq.), $MgSO_4$ (0.045 mmol, 10 eq.) and $MgCl_2$ (0.025 mmol, 5 eq.) in water (1 mL)*

*[b] Ribo-**30** (0.066 mmol, 13 eq.), 4-amino-2-methoxypyrimidine **180** (0.005 mmol, 1 eq.), $MgSO_4$ (0.045 mmol, 10 eq.) and $MgCl_2$ (0.025 mmol, 5 eq.) in water (1 mL)*

Attempted glycosylation of 4-amino-2-(methylthio)pyrimidine 181 with *ribo*-30

Entry	Initial pH	After 2 h 100 °C			After 4 h 100 °C		
		pH	**181**	**182**	pH	**181**	**182**
1[a]	7.1	5.2	100%	0%	5.0	100%	0%
2[a]	6.4	4.8	100%	0%	4.9	100%	0%
3[a]	6.0	5.3	100%	0%	5.1	100%	0%
4[b]	5.6	4.6	67%	33%	4.8	69%	31%
5[b]	4.9	4.6	74%	26%	4.5	77%	23%
6[b]	4.5	4.3	85%	15%	4.3	86%	14%

Table 7.4.20: *Relative yields (by ^{1}H NMR integration) of products observed from attempted glycosylation of 4-amino-2-(methylthio)pyrimidine **181** with ribo-**30**, General Procedure C.*

*[a] Ribo-**30** (0.069 mmol, 14 eq.), 4-amino-2-(methylthio)pyrimidine **181** (0.005 mmol, 1 eq.), $MgSO_4$ (0.045 mmol, 10 eq.) and $MgCl_2$ (0.025 mmol, 5 eq.) in water (1 mL). No nucleoside products could be seen by ^{1}H NMR analysis.*

*[b] Ribo-**30** (0.066 mmol, 13 eq.), 4-amino-2-(methylthio)pyrimidine **181**, (0.005 mmol, 1 eq.), $MgSO_4$ (0.045 mmol, 10 eq.) and $MgCl_2$ (0.025 mmol, 5 eq.) in water (1 mL). A new product **182** (δ 6.63 (d, J = 7.3 Hz, 1H) and 7.98 (br s, 1H)) was seen in all reactions (**Figure 7.4.24**).*

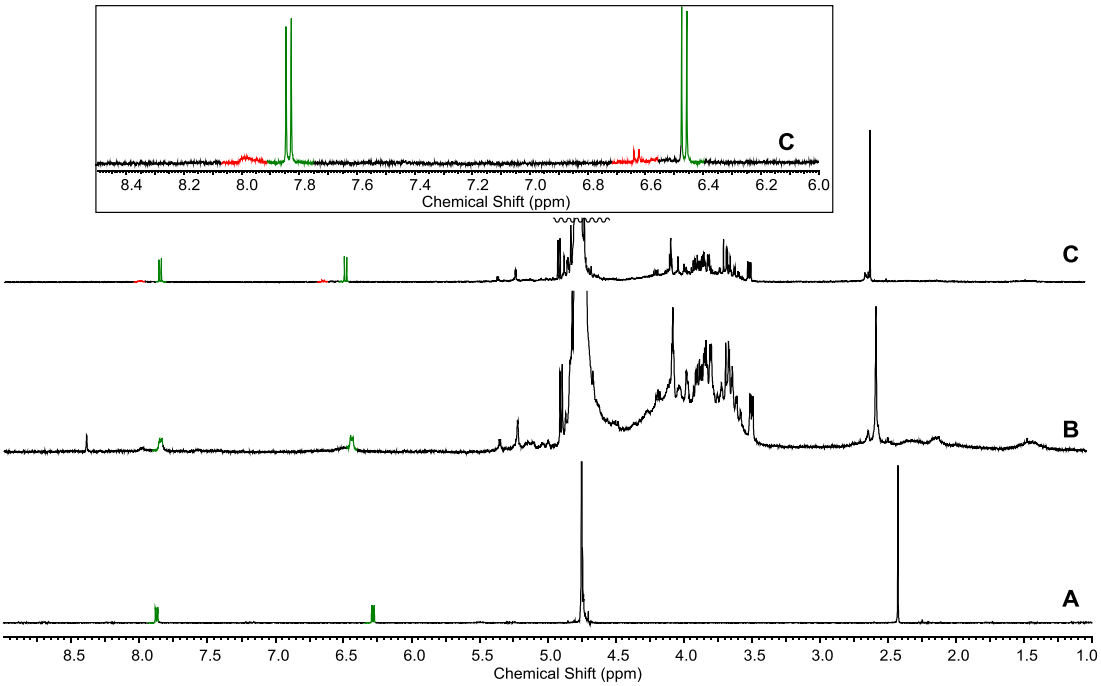


Figure 7.4.19 *^{1}H NMR spectra (600 MHz, D_2O, 1.0-9.0 ppm; inset 6.0-8.5 ppm) showing attempted glycosidation of 4-amino-2-(methylthio)pyrimidine **181** with ribo-**30**. **A**) **181** (1 eq.); and after heating with $MgSO_4$ (10 eq.) and $MgCl_2$ (5 eq.) at 100°C for 4 h: **B**) with ribo-**30** (14 eq.), initial pH 7.1 (**Table 7.4.10** entry 1); **C**) with ribo-**30** (13 eq.), initial pH 4.5 (**Table 7.4.10** entry 6). The peak coloured red at δ 6.63 (d, J = 7.3 Hz) integrates 1:1 with the broad peak at δ 7.98.*

Decomposition of 4-amino-2-methoxypyrimidine 180 under glycosylation conditions

4-Amino-2-methoxypyrimidine **180** (0.005 mmol, 1 eq.), $MgSO_4$ (0.045 mmol, 10 eq.) and $MgCl_2$ (0.025 mmol, 5 eq.) were dissolved in water (800 µL) and the solution adjusted to the desired pH, general procedure 4.3. ***Table 7.4.21***

Initial pH	After first period of heating at 100 °C				After second period of heating at 100 °C			
	Time / h	pH	**180**	**31**	Time / h	pH	**180**	**31**
6.0	4	10.5	7%	93%	8	10.5	0%	100%
5.5	2	10.6	21%	79%	-	-	-	-

Table 7.4.21: *Relative yields (by ^{1}H NMR integration) of 4-amino-2-methoxypyrimidine **180** and cytosine **31** observed during decomposition under glycosylation conditions of **180**.*

4-Amino-2-methoxypyrimidine **180** (0.005 mmol, 1 eq.), $MgSO_4$ (0.046 mmol, 9 eq.), $MgCl_2$ (0.025 mmol, 5 eq.) were dissolved in water (800 µL) and the solution was adjusted to the desired pH, general procedure 4.3. ***Table 7.4.22***

Initial pH	After 2 h 100 °C		
	pH	**180**	**3**
5.5	10.5	6%	94%
5.0	10.3	5%	95%

Table 7.4.22: *Relative yields (by ^{1}H NMR integration) of 4-amino-2-methoxypyrimidine* ***180*** *and cytosine* ***3*** *observed during decomposition under glycosylation conditions of* ***180.***

Decomposition of 4-amino-2-(methylthio)pyrimidine 181 under glycosylation conditions

4-Amino-2-(methylthio)pyrimidine **181** (0.005 mmol, 1 eq.), $MgSO_4$ (0.045 mmol, 10 eq.), $MgCl_2$ (0.025 mmol, 5 eq.) were dissolved in water (800 µL) and the solution adjusted to the desired pH, general procedure 4.3. ***Table 7.4.23.*** No other species could be observed by ^{1}H NMR analysis after heating in any sample.

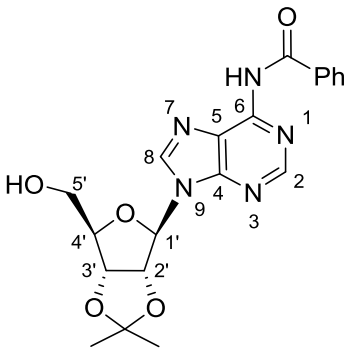
Entry	Initial pH	Heat time	Final pH
1	6.0	8 h	9.9
2	5.5	2 h	6.2
3	5.4	4 h	6.6
4	5.1	4 h	5.9

Table 7.4.23: *Decomposition of 4-amino-2-methoxypyrimidine* ***180*** *under glycosylation conditions.*

7.5 Experimental section for stereospecific reduction of 8,5′-cyclo purines

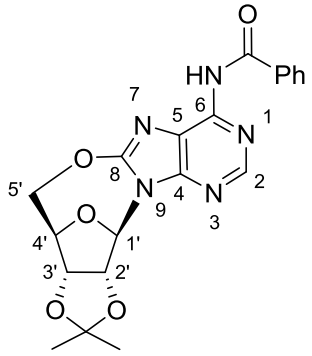
7.5.1 Synthesis and reduction of 2′,3′-*O,O*-isopropylidene-6-*N*-benzoyl-8,5′-anhydro-8-oxyadenosine *203*

2′,3′-O,O-Isopropyliene -6-N-benzoyladenosine **202b**



The method of van Delft *et al*. was used.[404] 2′,3′-O-isopropylidine adenosine **201** (299 mg, 0.97 mmol),was co-evaporated with pyridine, then dissolved in pyridine (5 mL, 0.19 mM). Trimethylsilyl chloride (0.65 mL, 5.12 mmol) was added and the solution stirred for 30 min at room temperature, followed by the addition of benzoyl chloride (0.15 mL, 1.29 mmol). After stirring for 3 h at room temperature, the solution was cooled to 0 °C and diluted with water (1 mL). After stirring for 10 min at 0°C, $NH_3 \cdot H_2O$ (2 mL) was added, and then stirred for a further 30 min at room temperature. The solvent was then removed by evaporation. The crude product was purified by column chromatography (hexanes:EtOAc 10:90 to EtOAc:MeOH 90:10) to give **202b** as a white powder (260 mg, 62%). ^{1}H NMR (600 MHz, $CDCl_3$) δ 8.80 (1H, s, H-2), 8.29 (1H, s, H-8), 8.04 (2H, d, *J* = 7.9 Hz, *o*H-Ph), 7.62 (1H, t, *J* = 7.2 Hz, *p*H-Ph), 7.53 (2H, t, *J* = 8.3 Hz, *m*H-Ph), 6.00 (1H, d, *J* = 3.8 Hz, H-1′), 5.24 (1H, br s, H-2′), 5.13 (1H, d, *J* = 4.9 Hz, H-3′), 4.58 (1H, s, H-4′), 4.01 (1H, d, *J* = 12.8 Hz, H-5′a), 3.83 (1H, d, *J* = 12.4 Hz, H-5′b), 1.66 (3H, s, CH_3), 1.40 (3H, s, CH_3). ^{13}C NMR (150 MHz, $CDCl_3$) δ 164.84, 152.57, 142.66 (C-8), 133.25, 133.21 (overlapping)(*p*C-Ph), 129.05 (*m*C-Ph), 128.16 (*o*C-Ph), 114.34 ($C(CH_3)_2$), 94.47 (C-1′), 86.78 (C-4′), 83.58 (C-2′), 81.87 (C-3′), 63.29 (C-5′), 27.62 (CH_3), 25.27 (CH_3). R_f 0.42 (EtOAc). M.p. > 114 °C (decomposes). IR (cm^{-1}) 3400-2090 (O-H, N-H), 1706 (C=O), 1603 (N-H). HRMS (*m/z*): $[M+H^+]^+$ calcd for formula $C_{20}H_{22}N_5O_5^+$, 412.1762; found 412.1761.

2′,3′-O,O-Isopropyliene-6-N-benzoyl-8,5′-anhydro-8-oxyadenosine **203**



The procedure of Maki *et al*. was used.[378] 2′,3′-O,O-Isopropyliene -6-N-benzoyladenosine **202b** (55.6 mg, 0.13 mmol), *N*-iodisuccinamide (90.0 mg, 0.4 mmol) and acetic acid (3 ml) were stirred at room temperature for 24 h. The solvent was then removed by evaporation. The crude product was purified by column chromatography (hexanes:EtOAc 50:50 to EtOAc:MeOH 50:50) and then recrystallised (EtOH) to give **203** as a white powder (30.55 mg, 58%). ^{1}H NMR (600 MHz, $CDCl_3$) δ 8.79 (1H, br s, H-6-N), 8.74 (1H, s, H-2), 7.99 (2H, d, *J* = 7.5 Hz, *o*H-Ph), 7.60 (1H, t, *J* = 7.5 Hz, *p*H-Ph), 7.51 (2H, t, *J* = 7.5 Hz, *m*H-Ph), 6.51 (1H, s, H-1′), 5.12 (1H, d, *J* = 5.3 Hz, H-2′), 4.79 (1H, d, *J* = 5.3 Hz, H-3′), 4.76 (1H, s, H-4′), 4.57 (1H, d, *J* = 12.8 Hz, H-5′a), 4.28 (1H, d, *J* = 12.8 Hz, H-5′b), 1.58 (3H, s, CH_3), 1.37 (3H, s, CH_3). ^{13}C NMR (150 MHz, $CDCl_3$) δ 164.60, 155.64, 152.20 (C-2), 150.71, 147.79, 133.64, 132.87 (*p*C-Ph), 128.97 (*m*C-Ph), 127.93 (*o*C-Ph), 119.76, 113.66 ($C(CH_3)_2$), 86.99 (C-1′), 85.91 (C-4′), 85.79, 81.32, 74.91 (C-5′), 26.19 (CH_3), 24.73 (CH_3). R_f 0.61 (EtOAc). M.p. > 112 °C (decomposes). IR (cm^{-1}) 3300-2090 (O-H, N-H), 1690 (C=O), 1600 (N-H). HRMS (*m/z*): $[M+H^+]^+$ calcd for formula $C_{20}H_{20}N_5O_5^+$, 410.1465; found 410.1466.

Reduction of ***203*** *to give* ***202b***

The procedure of Maki *et al.* was used.[369]

*Initial attempt***:** 2′,3′-*O,O*-Isopropyliene-6-*N*-benzoyl-8,5′-anhydro-8-oxyadenosine **203** (9.8 mg, 0.02 mmol), $NaCNBH_3$ (4.0 mg, 0.06 mmol) and AcOH (1 mL) were stirred at room temperature (approximately 15 °C) for 24 h. Water (0.5 mL) was added and stirred for a further 30 mins. Solvent was removed *in vacuo.* ^{1}H NMR integral analysis indicated that approximately 30% of **203** had converted to **202b**. The crude product was stirred with additional $NaCNBH_3$ (4.8 mg, 0.8 mmol) and AcOH (1 mL) at 25 °C for a further 25 h. Water (0.5 mL) was added and stirred for a further 30 mins. Solvent was removed *in vacuo.* ^{1}H NMR integral analysis indicated that approximately 70% of **203** had converted to **202b.** ^{1}H NMR spectrum shown in ***Figure 7.5.1.***

General procedure 5.1 *for reduction of nucleosides*:
Nucleoside, $NaCNBH_3$ and solvent (2 mL) were stirred at room temperature. After 65 h, additional $NaCNBH_3$ was added. After a further 24 h, water (0.5 mL) was added and stirred for a further 30 mins. Solvent was removed *in vacuo.*

Further attempted reductions of ***201*** *to* ***199*** were made using general procedure 5.1**.**

	203	**initial $NaCNBH_3$**	**additional $NaCNBH_3$**	**solvent**
1	18.2 mg, 0.05 mmol	7.3 mg, 0.11 mmol	6.4 mg, 0.10 mmol	AcOH
2	19.7 mg, 0.05 mmol	7.0 mg, 0.11 mmol	7.8 mg, 0.12 mmol	formic acid
3	18.3 mg, 0.05 mmol	–	-	formic acid
4	20.1 mg, 0.05 mmol	7.0 mg, 0.11 mmol	8.6 mg, 0.14 mmol	TFA

Table 7.5.1: *attempted reductions of* ***201*** *using $NaCNBH_3$. Entry 1 gave a single nucleotide, tentatively assigned as hydrolysis product* ***236 (***^{1}*H NMR spectrum shown in* ***Figure 7.5.1.)****; entries 2-4 gave a complex mixture of products.*

236: ^{1}H NMR (300 MHz, $CDCl_3$) δ 8.40 (1H, s, H-2), 8.03 (2H, d, *J* = 7.2 Hz, *o*-Ph), 7.67 (1H, t, *J* = 7.7 Hz, *p*-Ph), 7.57 (1H, t, *J* = 7.5 Hz, *m*-Ph), 6.34 (1H, d, *J* = 1.5 Hz, H-1′), 5.50 (1H, dd, *J* = 6.2, 1.3 Hz, H-2′), 5.07 (1H, dd, *J* = 6.2, 3.0 Hz, H-3′), 3.45 (1H, app t, *J* = 9.6 Hz, H-5′b) 3.27 (1H, dd, *J* = 9.6, 5.7 Hz, H-5′b).

Attempted reduction in $MeOH/HCl_{(aq)}$: 2′,3′-*O,O*-Isopropyliene-6-*N*-benzoyl-8,5′-anhydro-8-oxyadenosine **203** (21.5 mg, 0.05 mmol) and $NaCNBH_3$ (7.5 mg, 0.12 mmol) were added to a mixture of MeOH and 0.3 µM HCl (2 mL, 1:1 V/V, pH 3.7) and stirred at room temperature (approximately 15 °C) for 28 h. Solvent was removed *in vacuo.* ^{1}H NMR integral analysis (by relative integration of anomeric protons) indicated that approximately 56% of **203** remained, 35% had converted to a species tentatively assigned as **236** and 9% to a species assigned as **202b**. ^{1}H NMR spectrum shown in ***Figure 7.5.1.***

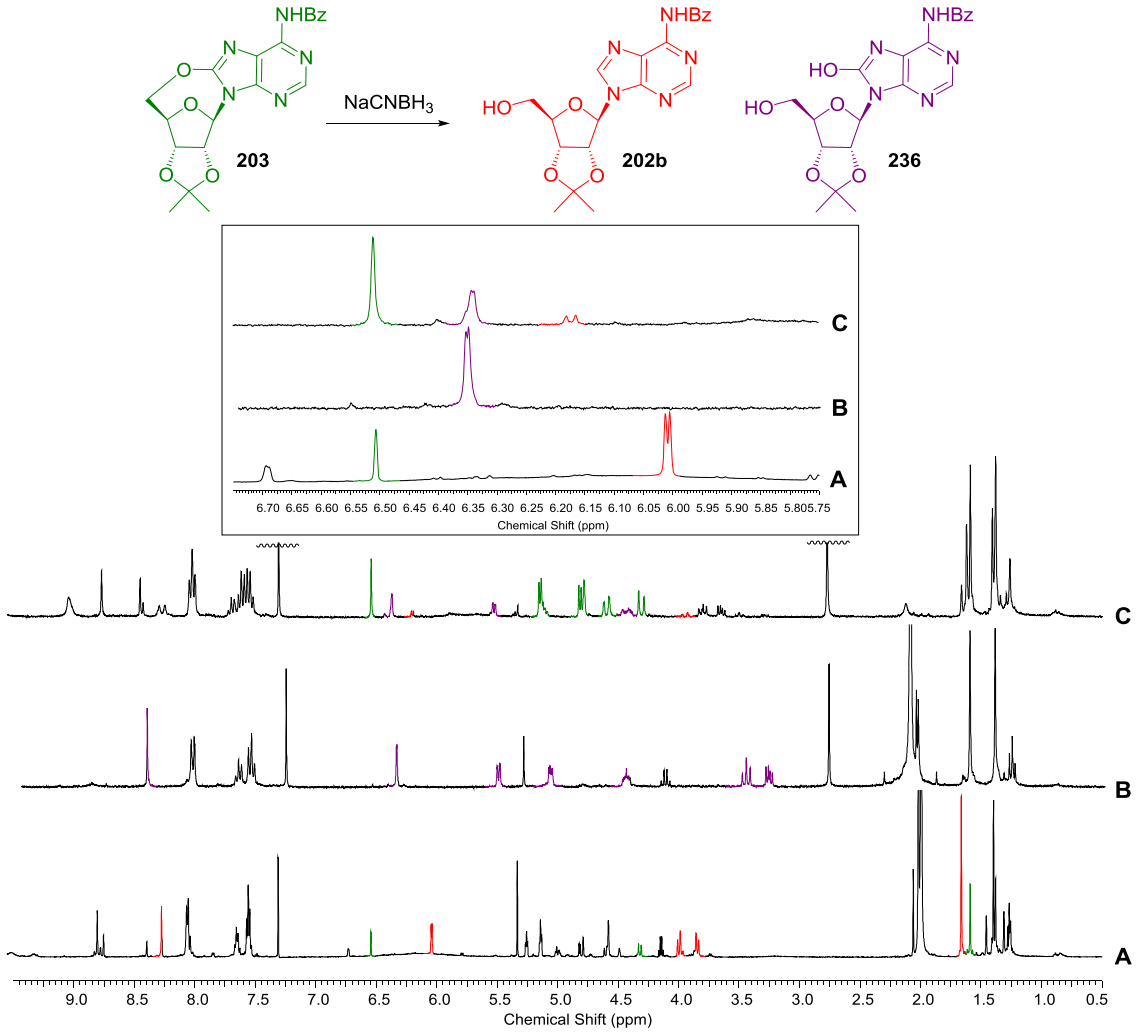
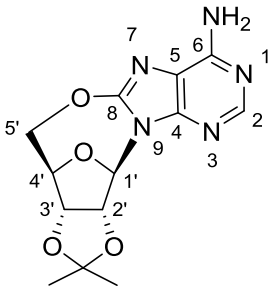


Figure 7.5.1: *^{1}H NMR spectra (300 MHz, $CDCl_3$, 0.5-9.5 ppm) of the crude products, after aqueous workup, of various attempted reductions of **203** (green) using $NaCNBH_3$, giving the reduction product **202b** (red) and the hydrolysis product **236** (purple), with the H-1' regions of each spectrum magnified (inset). **A)** $NaCNBH_3$ (7 eq.), AcOH after 25 h at r.t. Only **203** (30%, green) and **202b** (70%, red) are present. **B)** $NaCNBH_3$ (4 eq.), AcOH after 89 h at r.t. **203** has undergone complete conversion to a single nucleoside, believed to be the hydrolysis product **236** (in purple). The upfield shift of 5'-Hs and the resolution of 1'-H into a small doublet (δ 6.36, J = 1.5 Hz) indicates that the anhydro-ring was broken, but the absence of an H-8 peak indicates this position is substituted. Literature NMR data for **236** is not available (2',3'-O,O-isopropylidene-8-hydroxy adenosine 1'-H: δ 6.05, d, J = 5 Hz).[387] **C)** $NaCNBH_3$ (2.4 eq.), MeOH/ 0.3 μM HCl after 28 h at r.t. **203** (green) has undergone incomplete conversion to two nucleoside products: **236** (purple, 35%) and **202b** (red, 9%). Partial ^{1}H NMR data for **236**: δ 6.34 (1H, d J = 1.3 Hz, H-1'), 3.78 (1H, m, H-5'a), 3.63 (1H, m, H-5'b). Partial ^{1}H NMR data for **202b**: δ 6.17 (1H, d, J = 4.5 Hz).This coupling constant for this doublet is the same as for the desired product; small differences in pH can account for the difference in chemical shift.*

7.5.2 Synthesis of 2′,3′-*O*,*O*-isopropylidene-8,5′-anhydro-8-oxyadenosine *198*

2′,3′-O,O-Isopropylidene-8,5′-anhydro-8-oxyadenosine **198**



2′,3′-*O,O*-isopropylidine adenosine **201** (307 mg, 1.00 mmol) was stirred in a mixture of water (24 mL) and 1M NaOAc buffer (pH 4, 18 mL). Bromine water (7 ml, 1.50 mmol) was added and the resulting suspension stirred in the dark for 22 h. Additional bromine water was added at 22 h (5 mL, 1.1 mmol), 28 h (2 mL, 0.43 mmol), and 43 h (2 mL, 0.43 mmol). After a total of 52 h stirring, the addition of 4M NaOH raised the pH to 12.5. The solution was then heated at 60 °C for a further 96 h, with periodic additions of 1M NaOH to maintain the pH of the solution between pH 10 and pH 12. Compressed air was blown over the solution overnight to concentrate the solution by approximately half. The solution was extracted with $CHCl_3$ (3 x 15 mL) and concentrated *in vacuo* to give crude **30**. The purified product was obtained by recrystallization in hot EtOH to give **30** (0.18 g, 60%) as a white crystalline powder. ^{1}H NMR (600 MHz, DMSO) δ 8.12 (1H, s, H-2), 7.13 (2H, br s, NH_2), 6.03 (1H, s, H-1′), 5.11 (1H, d, *J* = 5.7, H-2′), 4.91 (1H, d, *J* = 5.6, H-3′), 4.75 (1H, s, H-4′), 4.64 (1H, dd, *J* = 12.9, 1.8 Hz, H-5′a), 4.14 (1H, d, J = 12.9 Hz, H-5′b), 1.47 (3H, s, CH_3), 1.30 (3H, s, CH_3). ^{13}C NMR (150 MHz, DMSO) δ 154.77 (C-6), 153.06 (C-8), 152.05 (C-2), 147.58 (C-4), 114.40 (C-5), 111.88 ($C(CH_3)_2$), 85.80 (C-1′), 85.21 (C-4′), 84.80 (C-3′), 80.90 (C-2′), 74.30 (C-5′), 25.91 (CH_3), 24.30 (CH_3). R_f 0.12 (12:1 $CHCl_3$:MeOH). M.p. > 228 °C (decomposes). IR (cm^{-1}) 3400-3000 (O-H, N-H), 1692 (N-H). HRMS (*m/z*): $[M+H^+]^+$ calcd for formula $C_{13}H_{16}N_5O_4^+$, 306.2102; found 306.1205. Data in agreement with literature.[387,381,389,392,405]

Attempted brominations under basic conditions: [380,382 dd]

Attempt A: 2′,3′-*O*,*O*-isopropylidine adenosine **201** (252 mg, 0.82 mmol), 10% Na_2HPO_4 aqueous solution (15 mL), dioxane (15 mL), and bromine (1.9 g, 12.1 mmol)

[dd] Although Ikehara *et al.* use a basic phosphate buffer in the brominations of isopropylidene-protected nucleosides,[380,390] they conversely conclude that pH 4 acetate buffer resulted in the highest yields in an extensive study on the effect of pH on the bromination of nucleotides.[406] A simple one-pot synthesis of **201** thus presented itself to us: if the bromination step could be carried out in aqueous solution at low pH, the cyclisation step could be triggered by raising the pH.

were stirred in darkness for 24 h. The mixture was then extracted with $CHCl_3$ (3 x 15 mL), washed with 1M $NaHSO_3$ (10 mL) and then water (10 mL), dried ($MgSO_4$), and concentrated *in vacuo* to give crude **204** in a complex mixture of products. Attempted recrystallisation from EtOH was unsuccessful.

Attempt B: 2′,3′-*O*,*O*-isopropylidine adenosine **201** (218 mg, 0.71 mmol) was dissolved in dioxane (20 mL). Na_2HPO_4 (2 g, 14 mmol) was dissolved in water (20 mL) to give a solution of pH 9.0, to which bromine (124 mg, 0.78 mmol) was added. The aqueous solution was added to the dioxane solution, and the mixture stirred in darkness for 25 h. The mixture was then extracted with $CHCl_3$ (3 x 15 mL), washed with 1M $NaHSO_3$ (10 mL) and then water (10 mL), dried ($MgSO_4$), and concentrated *in vacuo* to give crude **204** in a complex mixture of products. Flash column chromatography did not result in the isolation of **204**.

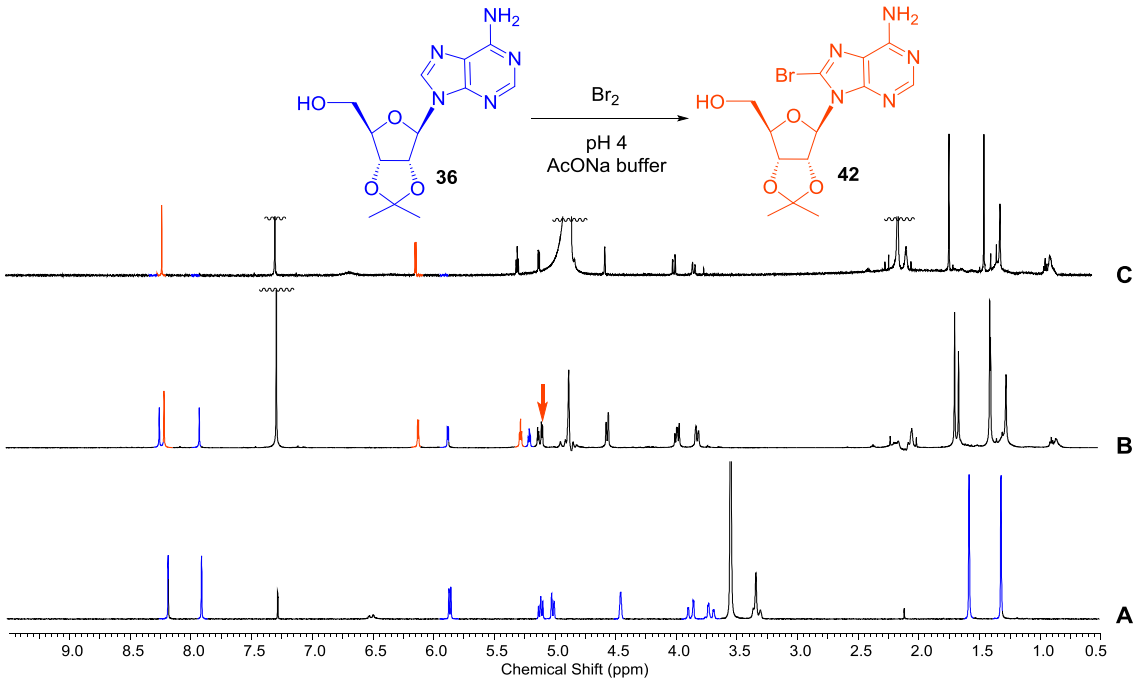


Figure 7.5.2: *Bromination of **201** in aqueous solution at pH 4. Reaction progress monitored by ^{1}H NMR (600 MHz, 0.5-9.5 ppm, $CDCl_3$ unless specified), after micro-extraction of aliquots of the reaction mixture into $CDCl_3$. **A)** Starting material **201** (in a mixture of $CDCl_3$ and CD_3OD); **B)** 22 h; **C)** 68 h.*

Reaction time / h	201 / %					204 / %				
	H-2	H-1'	H-2'	H-3'	mean	H-2	H-1'	H-2'	H-3'	mean
22	42	44	41	43	43	58	56	59	57	58
44	12	12	13	15	13	88	88	87	85	87
68	2	2	-	-	2	98	98	-	-	98

Table 7.5.2: *Yields of starting material and product over time, calculated by relative integrations of various peaks, where sufficiently distinguishable. (Reaction carried out using 108 mg, 0.35 mmol* ***201*** *in 26 mL solution)*

Initial pH	204 / %	198 / %
11.2	88	12
12.2	32	68
12.9	0	100

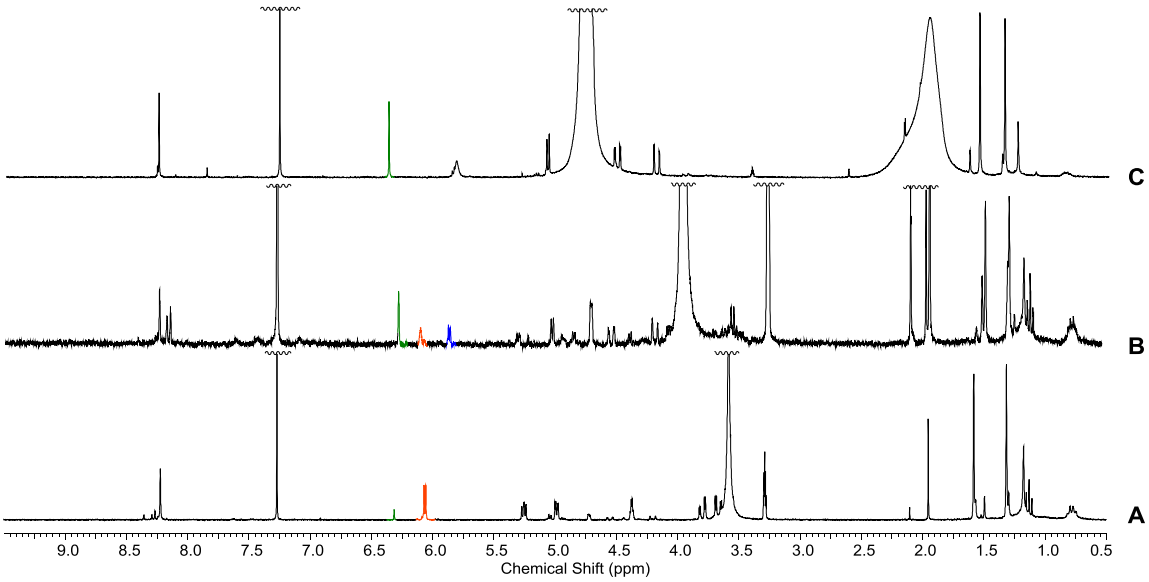


Figure 7.5.3: *Small scale studies of the progression of the cyclisation reaction by ^{1}H NMR analysis. 1mL aliquots of a completed bromination reaction (13 mM) were adjusted to various pHs with 1M NaOH and heated at 60°C for 5 h.* ***Top****: Table showing proportion of two species (by relative integration of anomeric protons) after heating at 60 °C for 5 h.* ***Bottom:*** *^{1}H NMR spectra (300 MHz, $CDCl_3$, 0.5-9.5 ppm) for the three initial pH values: A) pH 11.2; B) pH 12.2 (some starting material is also present: the ratio of the three species present by relative integration of anomeric protons is 49%* ***198****, 24%* ***204****, 27%* ***201****; C) pH 12.9.*

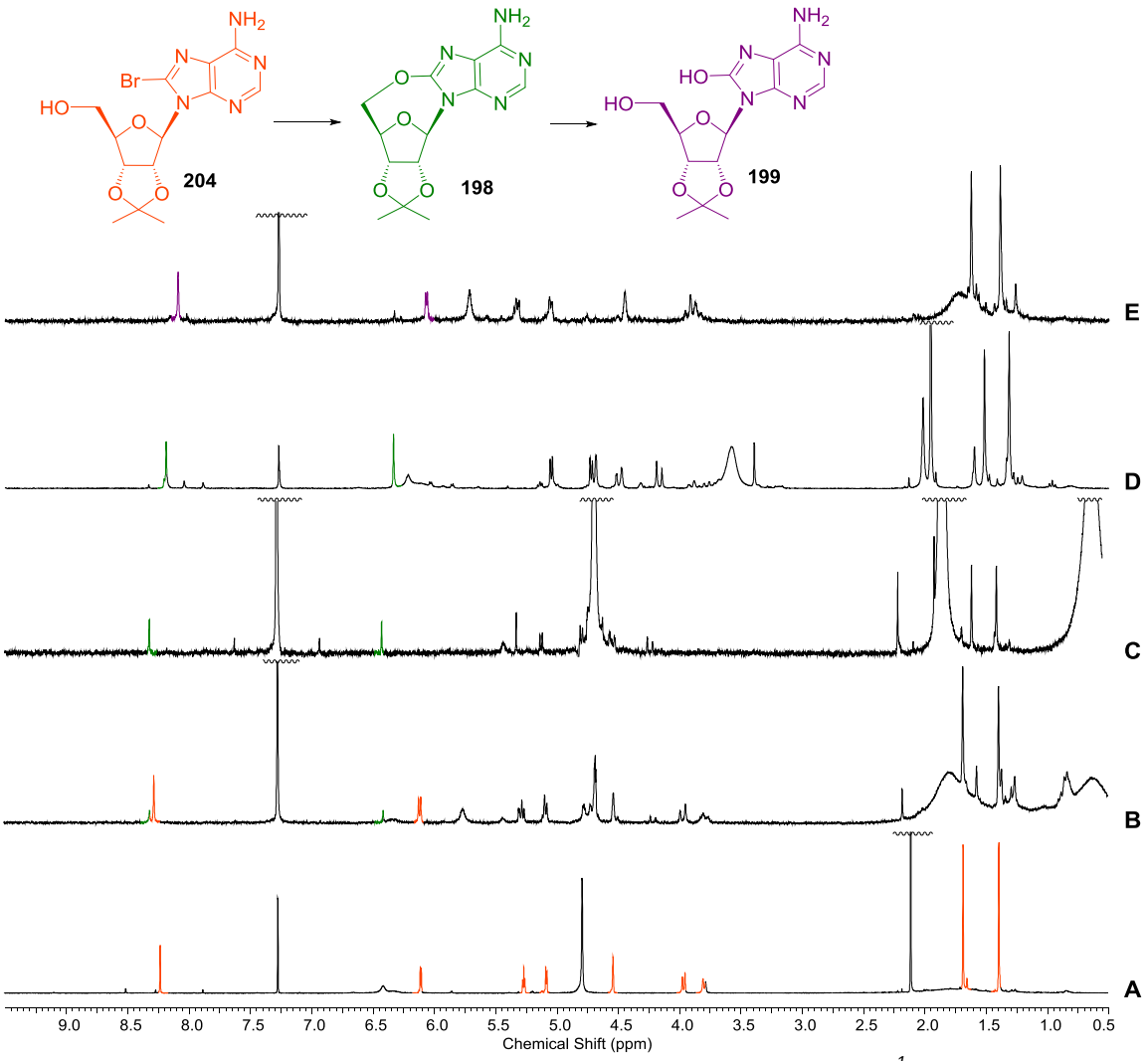
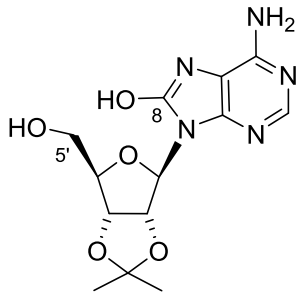


Figure 7.5.4: *Monitoring the progression of the cyclisation reaction by ^{1}H NMR. Reaction carried out on* ***201*** *(969 mg, 3.15 mmol) in 162 mL solution. The progression of the reaction was periodically monitored by removing aliquots of the reaction mixture and extraction into $CDCl_3$ for NMR analysis (300 MHz, $CDCl_3$, 0.5-9.5 ppm).* ***A)*** *Bromination appears complete after 48 h. The solution was adjusted from pH 3.5 to pH 12.9 and heated at 60 °C.* ***B)*** *After heating for 48 h at 60 °C, ^{1}H NMR analysis indicates starting material* ***204*** *and desired product* ***198*** *(23%) are the only nucleosides present. The solution was adjusted from pH 9.4 to 13.0.* ***C)*** *After a total of 72 h at 60 °C, ^{1}H NMR analysis indicates complete conversion to product* ***198.*** ***D)*** *The reaction mixture was extracted with $CHCl_3$, isolating crude* ***198*** *in just 11 % yield.* ***E)*** *The aqueous layer was adjusted to pH 6.4 and extracted as before to give crude* ***199*** *(25%).* ***199*** *H-1' ($CDCl_3$) δ 6.07 (d, J = 4.1 Hz);* ***199*** *H-1' (D_2O) δ 5.97 (d, J = 3.6 Hz).*

2′,3′-O,O-Isopropylidene -8-hydroxy adenosine **199**



Unintentional formation of **199** during attempted synthesis of **198:**

2′,3′-*O*,*O*-isopropylidine adenosine **201** (210 mg, 0.68 mmol), water (17 mL),1M NaOH (0.8 mL),1M NaOAc buffer (pH 4, 25 mL) and bromine water (7 mL. 0.5 mM, 3.5 mmol) were added and the resulting suspension stirred in the dark for 24 h. The pH was raised from 3.9 to 12.8 with the addition of 4 M NaOH (6mL) and the solutions heated at 60 °C for 26 h. After 26 h heating the pH was observed to be 12.6. The solution was extracted with $CHCl_3$ (5 x 25 mL), dried ($MgSO_4$), and concentrated *in vacuo* to give crude **198** (6.2 mg, 3%). An aliquot of the aqueous layer (400 µL) was removed and diluted with D_2O (100 µL): ^{1}H NMR analysis revealed the presence of one nucleoside product in solution, characterised by the presence of a doublet (*J* = 3.6 Hz) at δ 5.97. The aqueous layer was concentrated by half, and extracted with $CHCl_3$ (5 x 25 mL), dried ($MgSO_4$), and concentrated *in vacuo* to give crude **199** (12mg, 0.037 mmol, 5%). ^{1}H NMR (300 MHz, $CDCl_3$) δ 8.04 (1 H, s, H-2), 6.05 (1 H, d, *J* = 4.1 Hz, H-1′), 5.37 (1 H, dd, *J* = 3.96, 5.65, H-2′), 5.03 (1 H, dd, *J* = 1.32, 6.03, H-3′), 4.44 (1H, d, *J* = 1.9, H-4′), 3.92-3.88 (2H, m, H-5′), 1.59 (3H, s, CH_3), 1.37 (3H, s, CH_3). HRMS (*m/z*): $[M+H^+]^+$ calcd for formula $C_{13}H_{18}N_5O_5^+$, 324.1308; found 324.1301. Data in good agreement with literature.[387]

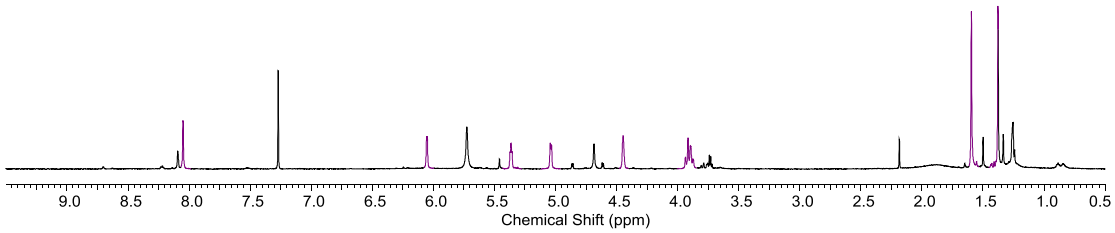


Figure 7.5.5: *^{1}H NMR spectrum (300 MHz, $CDCl_3$, 0.5-9.5 ppm) of* ***199.***

Second unintentional formation of **199** during attempted synthesis of **198:**

2′,3′-*O*,*O*-isopropylidine adenosine **201** (969 mg, 3.15 mmol), water (50 mL),1M NaOH (3.2 mL),1M NaOAc buffer (pH 4, 85 mL) and bromine water (24 mL, 77 mmol) were

added and the resulting suspension stirred in the dark for 30 h. Further bromine water was added (1 mL, 3.2 mmol) and stirred for a further 20 h. ^{1}H NMR analysis indicated that complete conversion ot 8-bromo had taken place. The pH was measured at this point as 3.47; it was raised to pH 12.94 with the addition of $NH_3 \cdot H_2O$ and NaOH. The solution was then heated at 60 °C for 50 h, at which point the pH was adjusted from 9.44 to 12.98. The solution was heated at 60 °C for a further 25 h. The solution was extracted with $CHCl_3$ (5 x 25 mL), dried ($MgSO_4$), and concentrated *in vacuo* to give crude **30** (0.11 g, 11%). The pH of the aqueous layer was raised to 6.4 and extracted as before to give crude **199** (0.25 g, 25%). **199**: ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.99 (1 H, s, H-2), 6.04 (1 H, d, *J* = 3.8 Hz, H-1′), 5.33 (1 H, dd, *J* = 5.74, 4.80 Hz, H-2′), 5.00 (1 H, app d, *J* = 4.8 Hz, H-3′), 4.42 (1H, app br s, H-4′), 3.9-3.85 (2H, m, H-5′), 1.59 (3H, s, CH_3), 1.36 (3H, s, CH_3). ^{13}C NMR (150 MHz, $CDCl_3$) δ 152.7, 151.2 (C-2), 147.2, 146.2, 114.1 (C(CH3)2), 104.0, 88.5 (C-1′), 86.2 (C-4′), 82.2 (C-2′), 81.2 (C-3′), 62.9 (C-5′), 27.5 (CH_3), 25.5 (CH_3). LRMS (*m/z*): $[M+H^+]^+$ calcd for formula $C_{13}H_{18}N_5O_5^+$, 324.32; found 324.11. HRMS (*m/z*): $[M+H^+]^+$ calcd for formula $C_{13}H_{18}N_5O_5^+$, 324.1308; found 324.1304. Data in good agreement with literature.[387]

*Deliberate synthesis of **199**:*

2′,3′-*O*,*O*-isopropylidine adenosine **201** (108 mg, 0.35 mmol), water (3 mL),1M NaOH (0.4 mL),1M NaOAc buffer (pH 4, 12.5 mL) and bromine water (10 mL, 1.94 mmol) were added and the resulting suspension stirred in the dark for a total of 68 h, with additional bromine water added after 22 h (5 mL, 0.97 mmol) and 44h (5 mL, 0.97 mmol). The reaction was monitered by periodic NMR analysis. The pH was raised from 4.1 to 12.7 with the addition of 4 M NaOH and the solutions heated at 60 °C for 4 d, after which time the pH was observed to be 13.0. The solution adjusted to pH 6 and was extracted with $CHCl_3$ (4 x 25 mL), dried ($MgSO_4$), and concentrated *in vacuo* to give crude **199** (1.68g) which was purified by flash column chromatography (EtOAc to EtOAc:MeOH 90:10) to give **199** as a white powder (57.0 mg, 0.18 mmol, 51%). ^{1}H NMR (600 MHz, $CDCl_3$) δ 8.02 (1 H, s, H-2), 6.04 (1 H, d, *J* = 3.8 Hz, H-1′), 5.33 (1 H, app t, *J* = 5.1 Hz, H-2′), 5.01 (1 H, dd, *J* = 6.1, 1.6 Hz, H-3′), 4.42 (1H, app br s, H-4′), 3.92 (1H, app d, *J* = 11.1 Hz, H-5′a), 3.86 (1H, dd, J = 12.33, 3.76 Hz, H-5′b), 1.58 (3H, s, CH_3), 1.35 (3H, s, CH_3).

Other syntheses of **198**

The procedure of Guo *et al.*[379] was attempted four times with various adaptions:

Entry	Concentration of *201*	Solvent	Eq. of NH_3	Eq. of I_2	Yield of *198*
1	250 mM	8 M NH_3/H_2O	24	4	1%
2	120 mM	3:5 8 M NH_3/H_2O:MeCN	24	6	-
3[a]	250 mM	8 M NH_3/H_2O	24	7	6%
4[a]	150 mM	7 M NH_3/MeOH	46	8	-

Table 7.5.3: *Adaptions of Guo et al.'s synthesis of 198 using iodine and ammonia. All took place at 60 °C with a reaction time of 24 h.* [a] Reaction took place in high-pressure vial

Table 7.5.3 *entry 1*: **201** (22.9 mg, 0.75 mmol) was heated to 60 °C with stirring in $NH_3 \cdot H_2O$ (3 mL, 8 M). Iodine (77.5 mg, 3.1 mmol) was added and the resulting suspension stirred for 24 h. Saturated aqueous Na_2SO_3 (5 mL) was added, and the crude product extracted into Et_2O (3 x 20 mL), washed with brine (5 mL), dried ($MgSO_4$) and solvent removed *in vacuo*. Purification by automated flash column chromatography ($CHCl_3$:MeOH 98:2 to 90:10) to give **198** (1.8 mg, 0.01 mmol, 1%).

Table 7.5.3 *entry 2*: **201** (30.2 mg, 0.98 mmol) was stirred in $NH_3 \cdot H_2O$ (3 mL, 8 M). The minimum amount of MeCN (5 mL) was added to the suspension until all starting material dissolved. The solution was heated to 60 °C and iodine (78.6 mg, 3.1 mmol) was added. Additional iodine (77.5 mg, 3.1 mmol) was added after 6.5 h. After a total of 24 h, saturated aqueous Na_2SO_3 (5 mL) was added to the solution, and MeCN removed *in vacuo* to give a suspension. A white precipitate was removed by filtration (identified as **32** by ^{1}H NMR analysis). The aqueous layer was extracted into Et_2O (3 x 20 mL), washed with brine (5 mL), dried ($MgSO_4$) and solvent removed *in vacuo*. ^{1}H NMR analysis revealed only **201** was present.

Table 7.5.3 *entry 3*: **201** (30.2 mg, 0.98 mmol), iodine (73.1 mg, 2.88 mmol) and $NH_3 \cdot H_2O$ (3 mL, 8 M) were placed in a high-pressure vial and stirred at 60 °C in the dark for 5 h. The suspension was then cooled, additional iodine (70.5 mg, 2.78 mmol) added, and again stirred at 60 °C in the dark for 15 h. The suspension was again cooled, iodine (39.3 mg, 1.55 mmol) added and stirred at 60 °C in the dark for 4 h. Saturated aqueous Na_2SO_3 (5 mL) was added to the suspension and a white precipitate was removed by filtration. ^{1}H NMR integral analysis of the precipitate indicated it was composed of approximately 25% of **198** with **201**. ^{1}H NMR analysis of the aqueous layer indicated it did not contain **198**. The precipitate was purified by automated flash column chromatography ($CHCl_3$:MeOH 98:2 to 90:10), the fractions

containing the desired product combined, the solvent removed *in vacuo* and recrystallized (EtOH) to give **198** (19.1 mg, 0.06 mmol, 6%).

Table 7.5.3 *entry 4*: **201** (30.2 mg, 0.98 mmol), iodine (75.7 mg, 2.98 mmol) and NH_3·MeOH (6.5 mL, 7 M) were placed in a high-pressure vial and stirred at 60 °C in the dark for 5 h. The solution was cooled and additional iodine (71.2 mg, 2.81 mmol) added, and again stirred at 60 °C in the dark for 15 h. The solution was again cooled, iodine (60.7 mg, 2.40 mmol) added and stirred at 60 °C in the dark for 4 h. Saturated aqueous Na_2SO_3 (5 mL) was added and the resulting solution concentrated *in vacuo* until a precipitate formed. The suspension was filtered (^{1}H NMR analysis of the precipitate showed that it did not contain **198**), the filtrate again concentrated *in vacuo* until a precipitate formed and filtered (^{1}H NMR analysis of this second precipitate showed that it was not **198**). The filtrate was extracted in Et_2O (3 x 10 mL), washed with brine (5 mL), and solvent removed *in vacuo*. The crude white solid was identified as **201** by ^{1}H NMR analysis.

The procedure of Maki *et al.*[378] was attempted twice:

201 (73.9 mg, 0.24 mmol) and *N*-iodosuccinamide (16.8 mg, 0.75 mmol) were stirred in 2 mL AcOH in the dark for 24 h. The solvent was removed *in vacuo.* ^{1}H NMR analysis of the crude showed only **201** was present.

201 (77.1 mg, 0.25 mmol) and NIS (16.6 mg, 0.74 mmol), AcOH (2 mL), 96 h. The solvent was removed *in vacuo.* ^{1}H NMR integral analysis of the crude showed approximately 10% conversion to **198**. Recrystallisation (EtOH) resulted in the isolation of only **201** and NIS.

7.5.3 Attempted reduction of 2′,3′-*O*,*O*-isopropylidene-8,5′-anhydro-8-oxyadenosine *198*

	198	**initial $NaCNBH_3$**	**additional $NaCNBH_3$**	**solvent**	**^{1}H NMR analysis after stirring for 60 h indicated:**
1	15.3 mg, 0.05 mmol	6.7 mg, 0.11 mmol	5.3 mg, 0.09 mmol	AcOH	no change had taken place
2	10.5 mg, 0.03 mmol	7.2 mg, 0.12 mmol	3.2 mg, 0.05 mmol	TFA	a complex mixture of products was present; negative spike with **201**

Table 7.5.4 *Attempted reduction of 2',3'-O,O-isopropylidene-8,5'-anhydro-8-oxyadenosine* ***198*** *using general procedure 5.1.* [369] *^{1}H NMR spectra shown in* ***Figure 7.5.6***

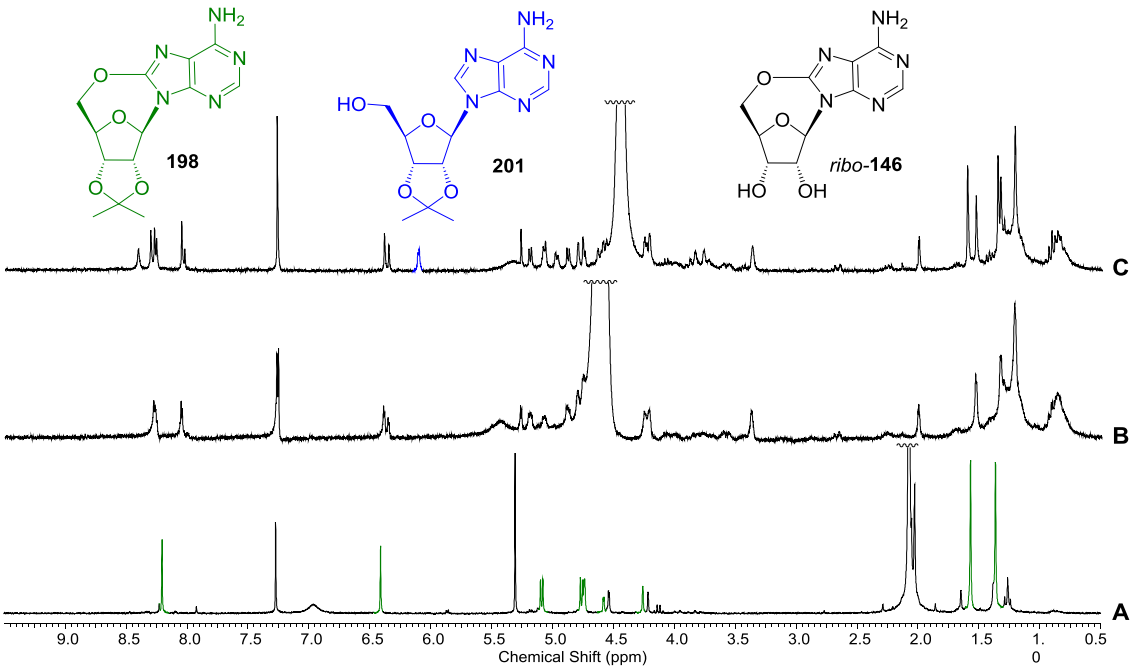


Figure 7.5.6: *Comparison of sections of ^{1}H NMR spectra (300 MHz, $CDCl_3$, 0.5-9.5 ppm) of attempted reductions of **198** with $NaCNBH_3$. **A**) $NaCNBH_3$ (4 eq.), AcOH, 60 h: no change **B**) $NaCNBH_3$ (5.7 eq.), TFA, 60 h: two species observed; **C**) negative spike with 2',3'-isopropylidine adenosine **201**. The unidentified product has a similar H-1' chemical shift to the starting material, indicating that it is strained. It could be ribo-**146**, but peaks in the isopropyl region cannot be accurately integrated to confirm this hypothesis.*

General procedure 5.2 *for reduction of nucleosides*:

2',3'-O,O-isopropylidene-8,5'-anhydro-8-oxyadenosine **198**, $NaCNBH_3$ and solvent (2 mL) were stirred at room temperature. At various timepoints, additional $NaCNBH_3$ was added. After a total of 69 h, water (0.5 mL) was added and stirred for 30 mins, then additional water (0.25 mL) added and stirred for a further 30 mins. Solvent was removed *in vacuo*, the sample dissolved in D_2O and analysed by ^{1}H NMR. The sample was re-dissolved in solvent (2 mL) and $NaCNBH_3$ again added. Further $NaCNBH_3$ was added at various timepoints. After a total of 133 h stirring (64 h since re-dissolving), water (0.5 mL) was added and stirred for 30 mins, then additional water (0.25 mL) added and stirred for a further 30 mins. Solvent was removed *in vacuo*, the sample dissolved in DMSO and analysed by ^{1}H NMR.

	Solvent		198	NaCNBH$_3$							
				0 h	28 h	47 h		69 h	86 h	93 h	109 h
1	AcOH	mg	20.3	11.4	12.1	17.4		15.5	16.3	23.8	37.5
		mmol	0.07	0.18	0.19	0.28		0.25	0.26	0.38	0.60
2	TFA	mg	19.9	12.4	12.7	17.9		17.7	15.8	22.3	34.5
		mmol	0.07	0.20	0.20	0.29		0.28	0.25	0.36	0.55

Table 7.5.5: *Attempted reduction of 2′,3′-O,O-isopropylidene-8,5′-anhydro-8-oxyadenosine* ***30*** *using general procedure 5.2.*

Table 7.5.5, *Entry 1:* After 69 h reaction time, ^{1}H NMR analysis showed only 4% conversion to an unidentified product. After 133 h reaction time, 35% conversion to the unidentified product had taken place. The two species can be distinguished by their aromatic and H-1′ signals; all other signals overlap. The unidentified product has a similar H-1′ (singlet) chemical shift to **198**, demonstrating its strained nature, Isopropylidene peaks were present, but overlap with impurity peaks preventing their integration. Following data assigned as far as possible for **198** and unidentified product (***Figure 7.5.7***).

198: ^{1}H NMR (600 MHz, D_2O) δ 8.17 (1H, s, H-2), 6.24 (1H, s, H-1′).

Unidentified product: ^{1}H NMR (600 MHz, D_2O) δ 8.38 (1H, s, H-2), 6.29 (1H, s, H-1′).

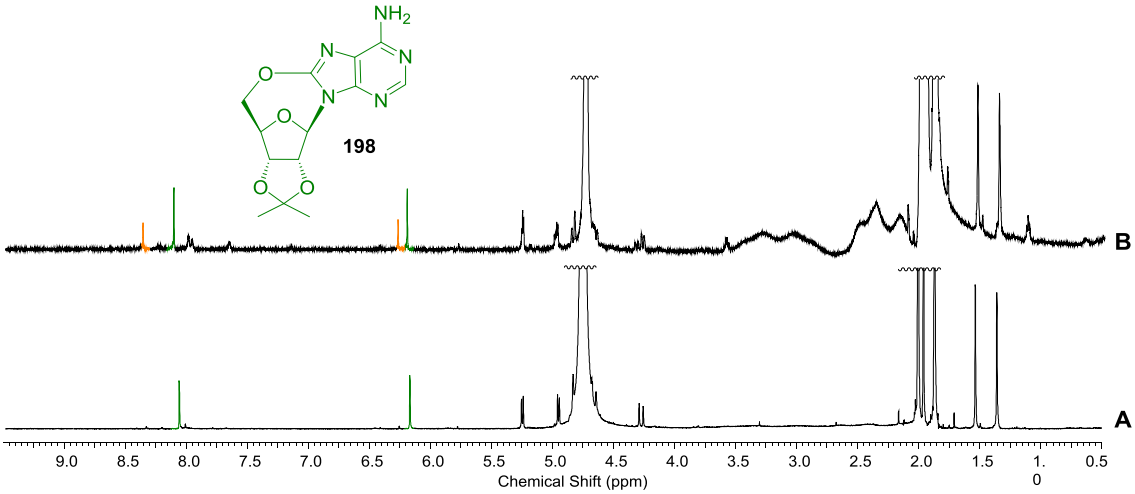


Figure 7.5.7: *Comparison of sections of crude ^{1}H NMR spectra (600 MHz, D_2O, 0.5-9.5 ppm) of attempted reduction of* ***198*** *with NaCNBH$_3$ in AcOH after:* ***A****) 69 h; and* ***B****) 133 h. A new nucleoside product is present, characterised by peaks highlighted in orange.*

Table 7.5.5, *Entry 2:* After 69 h reaction time, ^{1}H NMR analysis showed that no isopropylidene-protected species were present. Two species were present, characterised by a 1′ signals at δ 6.02 (s) and δ 5.62 (d, *J* = 5.3 Hz) in a 26:1 ratio (in

D_2O). After 133 h reaction time, the same species were present in a 5:3 ratio. Following data assigned as far as possible for major and minor products.

Major: ^{1}H NMR (600 MHz, DMSO) δ 8.24 (1H, s, H-2), 6.10 (1H, s, H-1′), 5.16 (1H, d, *J* = 5.8 Hz, H-2′), 4.99 (1H, d, *J* = 5.8 Hz, H-3′), 4.78 (1H, br s, H-4′), 4.67 (1H, d, *J* = 12.9 Hz, H-5′a), 4.15 (1H, d, *J* = 12.9 Hz, H-5′b).[e]

Minor: ^{1}H NMR (600 MHz, DMSO) δ δ 8.01 (1H, s, H-2), 5.67 (1H, d, *J* = 6.4 Hz, H-1′), 4.84 (1H, app t, *J* = 6.1 Hz, H-2′), 4.15 (1H, m, partially obscured by major H-5′b, H-3′); 3.88 (1H, app d, *J* = 3.7 Hz, H-4′), 3.60 (1H, ABX, *J* = 12.1, 3.7 Hz, H-5′a), 3.46 (1H, ABX, *J* = 12.3, 4.3 Hz, H-5′b).

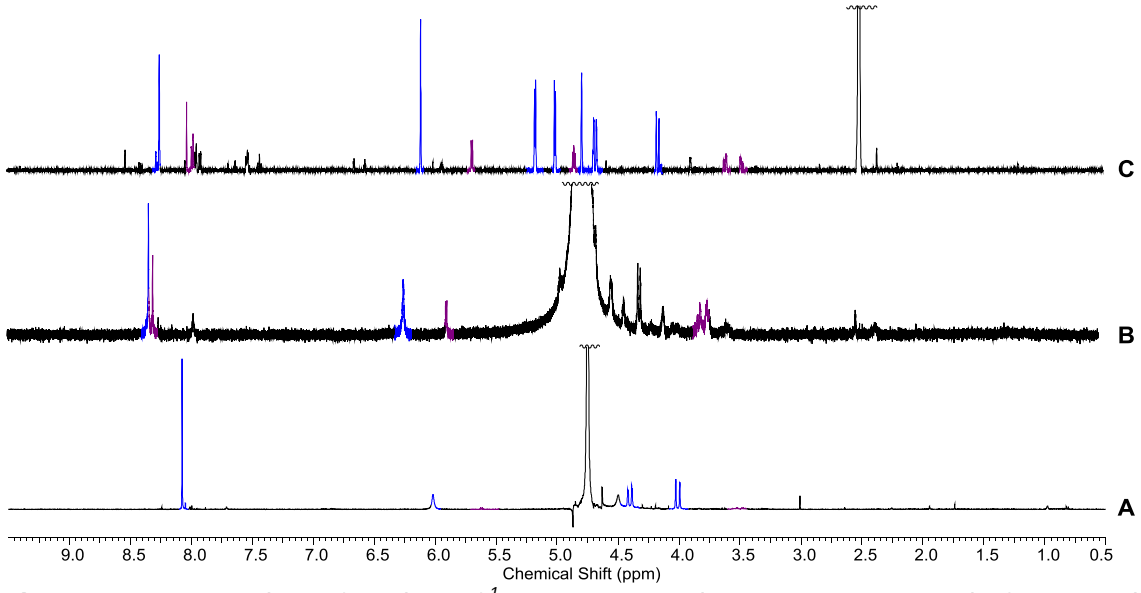


Figure 7.5.8: *Comparison of sections of ^{1}H NMR spectra (400 MHz, 0.5-9.5 ppm) of attempted reduction of* ***198*** *with $NaBH_3CN$ in TFA.* ***A****) after stirring for 69 h, the starting material has undergone complete conversion to two products, highlighted in blue and purple (D_2O);* ***B****) after 133 h in D_2O;* ***C****) same time point in DMSO: the H-2 peaks for the two species are better distinguished.*

[e] The chemical shifts and multiplicity for the anomeric and H-5′ protons indicate that a degree of ring strain is present in the major product, but comparison with literature ^{1}H NMR data for 5′,8-anhydro adenosine *ribo*-**146 (*Table 7.5.6*)** demonstrates that this is not present (*ribo*-**146** has distinctively well separated H-5′ chemical shifts; the presence of adjacent H-5′ protons at δ 3.60 and 3.46 in the product formed is corroborated by 2D NMR spectra).

Assignment	Yu[392] (400 MHz, DMSO)	Capon[393] (500 MHz, DMSO)	Ikehara[407] (MHz and solvent not stated)	Ikehara [381] (60 MHz, DMSO/D_2O)
H-2	8.12 (s, 1H)	8.12 (s)	8.10 (s)	8.2 (s)
6-NH_2	7.07 (s, 2H)	7.05 (br s)	7.04 (s)	
H-1′	5.98 (s, 1H)	5.98 (s)	6.05 (d, *J* = 0 Hz)	6.07 (s)
OH	5.65 (d, *J* = 6.4 Hz, 1H)	5.63 (br s)		
OH	5.35 (d, *J* = 5.2 Hz, 1H)	5.34 (br s)		
H-5′a	4.59 (d, *J* = 12.8 Hz, 1H)	4.59 (d, *J* = 12.9 Hz)		
H-4′	4.56 (s, 1H)	4.56 (s)		
H-3′	4.46 (dd, *J* = 5.6, 5.2 Hz, 1H)	4.46 (d, *J* = 6.1 Hz)		
H-2′	4.25 (dd, *J* = 5.6, 6.4 Hz, 1H)	4.25 (d, *J* = 6.0 Hz)		
H-5′b	4.09 (d, *J* = 12.8 Hz, 1H)	4.09 (d, *J* = 12.8 Hz)		

Table 7.5.6: *Literature ^{1}H NMR data for 5′,8-anhydro adenosine ribo-**146***

7.5.4 Attempted diol deprotection of 2′,3′-*O*,*O*-isopropylidene-8,5′-anhydro-8-oxyadenosine *198*

Initial attempt with H_2SO_4[381]

198 (19.8 mg, 0.06 mmol) was stirred in 0.5M H_2SO_4 (0.6 mL) at 60 °C for 5 h. The solution was neutralised with Dowex (OH^- form) and the solvent removed *in vacuo*. ^{1}H NMR integral analysis indicated a complex mixture of products was present, but analysis of the crude reaction mixture by mass spectrometry indicated the desired product *ribo*-**146** was present. HRMS ($[C_{10}H_{10}N_5O_4]^-$) calcd. 264.0733, found 264.5118. ***Figure 7.5.9***

Second attempt with H_2SO_4:

198 (32.5 mg, 0.11 mmol) was stirred in 0.5M H_2SO_4 (0.9 mL) at 60 °C for 4 h. The solution was neutralised with Dowex (OH^- form) and the solvent removed *in vacuo*. ^{1}H NMR analysis indicated only **198** was present. The sample was stirred again in 0.5M H_2SO_4 (0.9 mL) at 60 °C for 7 h, the solution neutralised and solvent removed *in vacuo*. ^{1}H NMR analysis indicated that no further change had occurred. (***Figure 7.5.10***)

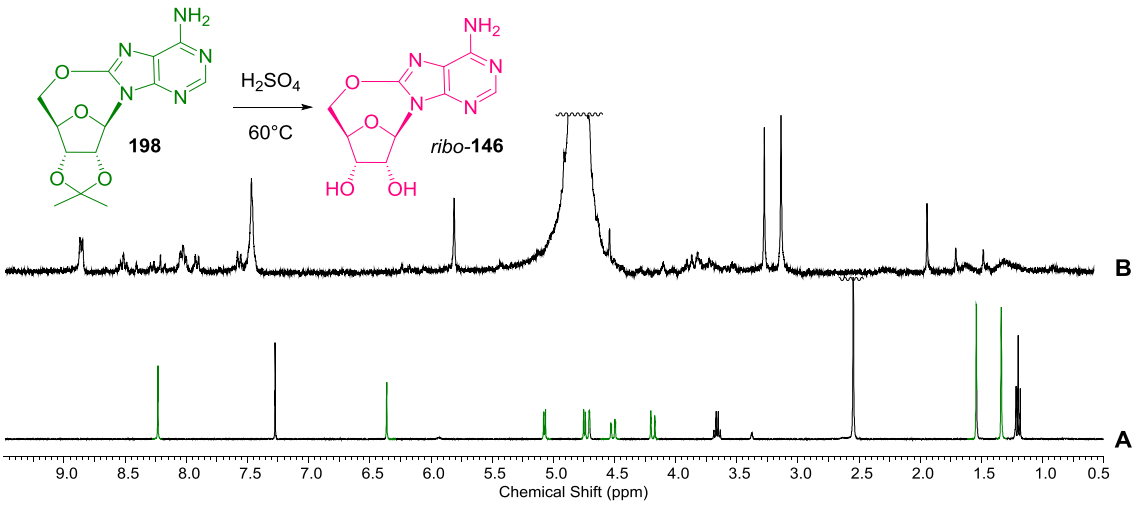


Figure 7.5.9: *^{1}H NMR spectra (0.5-9.5 ppm) showing attempted diol deprotection of **198** in H_2SO_4. **A**) **198** (600 MHz, $CDCl_3$); and **B**) complex mixture of products seen by ^{1}H NMR analysis (300 MHz, D_2O) after heating at 60 °C for 5 h.*

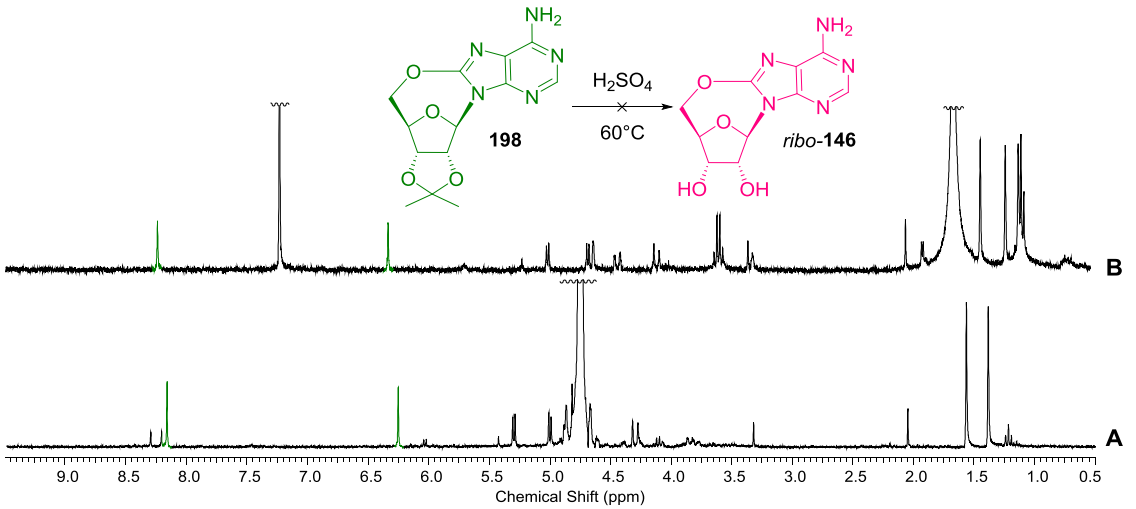


Figure 7.5.10: *^{1}H NMR spectra (300 MHz, 0.5-9.5 ppm) showing second attempted diol deprotection of **198** in H_2SO_4. After heating at 60 °C for: **A**) 4 h (in D_2O); and **B**) 11 h (in $CDCl_3$). **198** is the only nucleoside visible by ^{1}H NMR at each time point, and its H-2 and anomeric protons integrate to the isopropylidene peaks.*

Initial attempt with AcOH:

198 (19.4 mg, 0.06 mmol) was stirred in 0.75 M AcOH (0.9 mL) at 60 °C for 22 h, and then at 90 °C for a further 10 h. The solvent was removed *in vacuo*. A complex mixture of products formed;^{1}H NMR integral analysis indicated that partial deprotection had taken place (***Figure 7.5.11***).

^{1}H NMR data assigned as far as possible for four nucleoside species seen by ^{1}H NMR spectroscopy (yields based on relative integrations of anomeric protons) (300 MHz, D_2O):

198 (33%): δ 8.26 (1H, s, H-2), 6.29 (1H, s, H-1′), 5.27 (1H, d, *J* = 5.8 Hz, H-2′), 5.00 (1H, d, *J* = 5.5 Hz, H-3′), 1.53 (3H, s, CH_3), 1.36 (3H, s, CH_3).[f]

ribo-**146** (47%): δ 8.25 (1H, s, H-2), 6.19 (1H, s, H-1′).[g]

Product C (7%): δ 6.05 (1H, d, *J* = 6.4 Hz, H-1′).

Product D (13%): δ 5.82 (1H, d, *J* = 6.2 Hz, H-1′).[h]

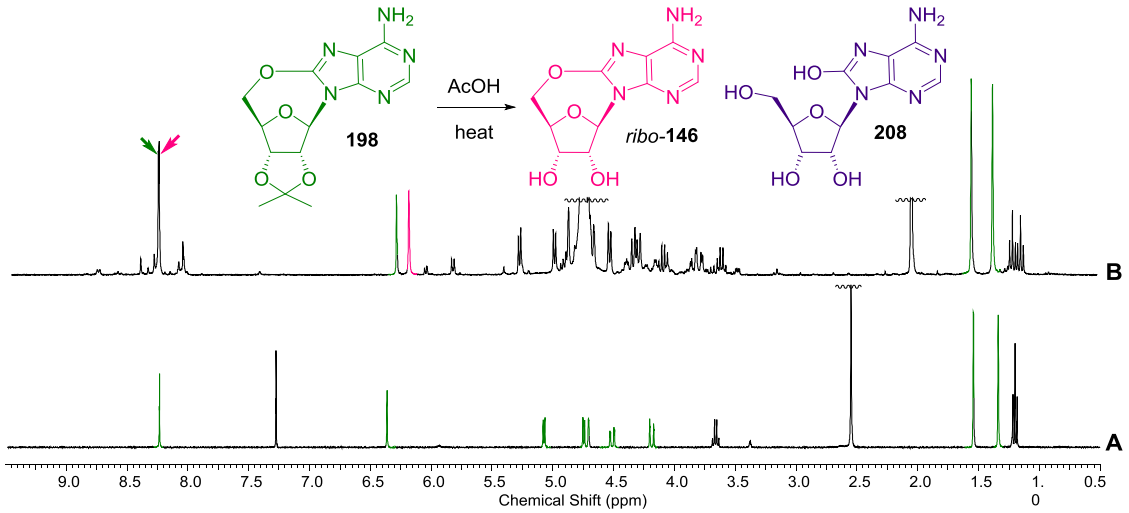


Figure 7.5.11: *^{1}H NMR spectra (0.5-9.5 ppm) showing partial deprotection of* ***198*** *(in green) in AcOH.* ***A)*** ***198*** *(600 MHz, $CDCl_3$);* ***B)*** *(300 MHz, D_2O) after heating in 0.75 M AcOH for 32 h (60 °C for 22 h, and then at 90 °C for a further 10 h) has undergone partial conversion to 5′,8-anhydro adenosine (ribo-****146,*** *pink) and two other species, both of which display H-1′ peaks as doublets indicating a lesser degree of ring strain. The isopropylidine peaks integrate to the starting material only. A plausible by-product (which is less strained than the starting material and desired product) is the hydrolysis product* ***208****.*

Second attempt with AcOH:

2′,3′-*O,O*-Isopropylidene-8,5′-anhydro-8-oxyadenosine **198** (32.9 mg, 0.11 mmol) was stirred in 1M AcOH (0.9 mL) at 90 °C for 7 h, then further 1M AcOH (0.9 mL) was added and the solution stirred for a further 18 h at 80 °C. The solvent removed *in*

[f] Only one set of isopropylidine peaks were apparent, which integrate exactly to the H-1′ peak attributed **198.**

[g] The lack of multiplicity implies that the ring strain conferred by the 8,5′-anhydro bond alone is sufficient to lock the anomeric proton in place, preventing it from coupling to H-2′.

[h] It is possible that product C or D is 8-oxo-adenosine **208,** the H-1′ peaks for which have been variously reported as: δ 5.68 (d, *J* = 6.1 Hz), 5.66 (d, *J* = 6.5 Hz), and 5.60 (1H, d, *J* = 2.0 Hz) in DMSO, 5.90 (d, *J* = 6.5 Hz, 1H) in D_2O, and 5.76 (d, *J* = 5.5) in H_2O.

vacuo. ^{1}H NMR analysis indicated only **198** was present. The sample was again stirred in 1M AcOH (0.9 mL) at 80 °C for 8 h and the solvent removed *in vacuo*. ^{1}H NMR analysis indicated that no further change had taken place. The sample was then stirred in 9M AcOH (0.9 mL) at 60 °C for 7 h and the solvent was removed *in vacuo*. ^{1}H NMR integral analysis indicated that no further change had taken place (***Figure 7.5.12***).

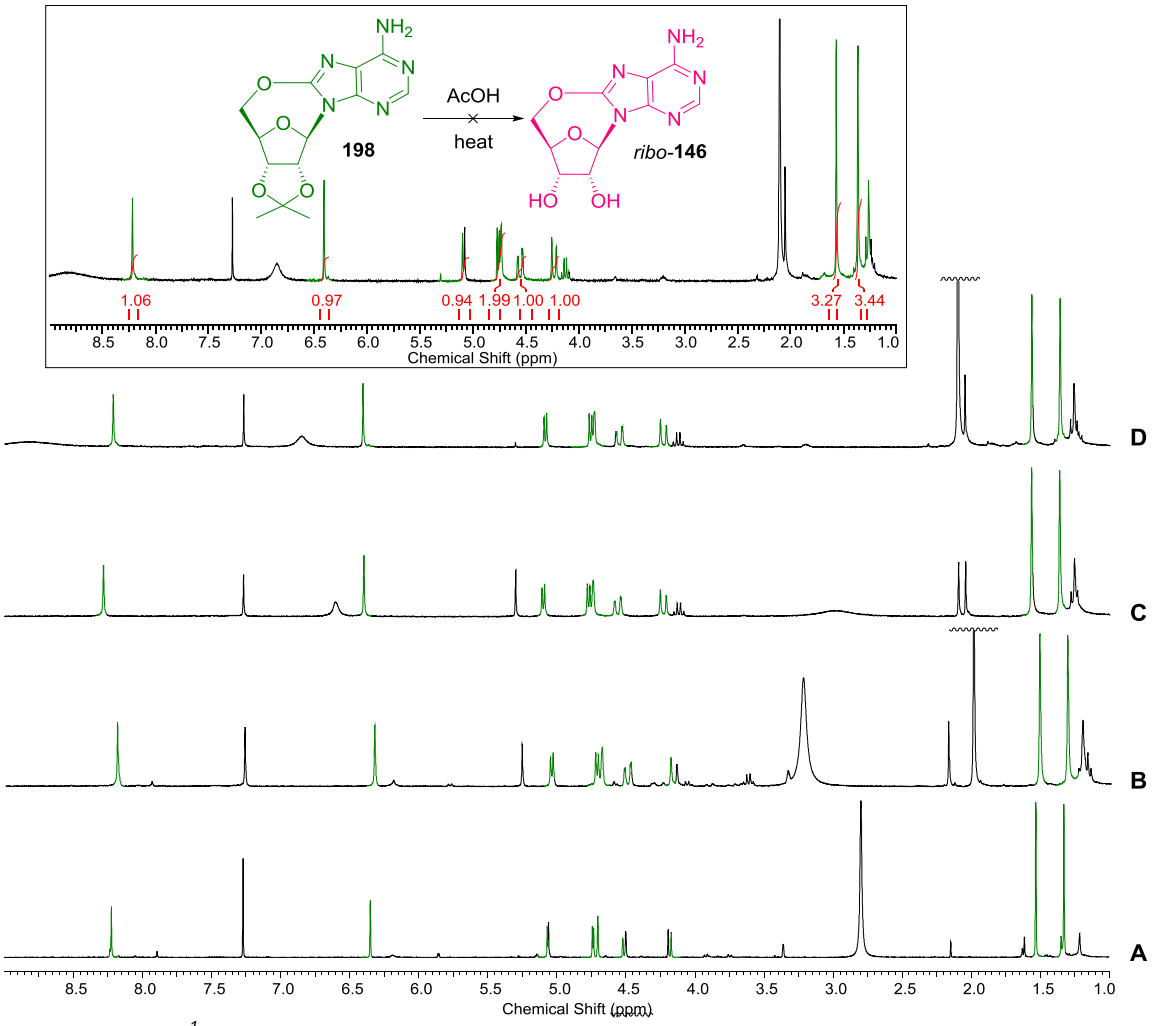
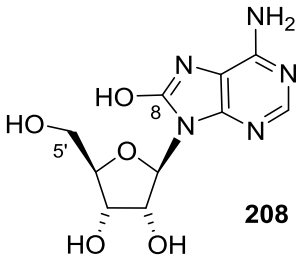


Figure 7.5.12: *^{1}H NMR spectra (300 MHz, $CDCl_3$, 1.0-9.0 ppm) showing second attempted deprotection of **198** (in green) in AcOH, after heating: **A**) at 90°C for 7 h; **B**) at 80°C for a further 18 h; **C**) at 80 °C for an additional 8 h; and **D**) at 60 °C for an additional 7 h. The size of the isopropylidine peaks have not decreased in comparison to the other nucleoside peaks.*

Attempt with TFA in MeOH/H_2O:

The procedure of Yu et al was used.[392] 2′,3′-*O,O*-Isopropylidene-8,5′-anhydro-8-oxyadenosine **198** (23.0 mg, 0.08 mmol) was refluxed in a mixture of TFA (1 mL), water (1 mL) and MeOH (2 mL) for 5 h. The solvent was removed *in vacuo*. ^{1}H NMR

analysis indicated only **198** was present. The sample was again refluxed in TFA (1 mL), water (1 mL) and MeOH (2 mL) for 18 h, and the solvent removed *in vacuo*. ^{1}H NMR analysis indicated that complete conversion to 8-oxo-adenosine **208** had taken place (***Figure 7.5.13***).



208

^{1}H NMR (600 MHz, DMSO) δ 7.98 (1H, s, H-2), 5.66 (1H, d, *J* = 6.6 Hz, H-1′), 5.30-5.25 (2H, m, 2′-OH, 5′-OH), 5.07 (1H, br s, 3′-OH), 4.87-4.84 (1H, m, H-2′), 4.13-4.11 (1H, m, H-3′), 3.87-3.85 (1H, m, H-4′), 3.62-3.58 (1H, m, H-5′a), 3.48-3.43 (1H, m, H-5′b). ^{1}H NMR (600 MHz, D_2O) δ 7.90 (1H, s, H-2), 5.80 (1H, d, *J* = 6.8 Hz, H-1′), 4.35 (1H, dd, *J* = 5.3, 2.2 Hz), 4.15 (1H, d, *J* = 2.4 Hz), 3.91 (1H, ABX, *J* = 13.1 Hz), 3.82 (1H, ABX, *J* = 13.7, 2.9 Hz). ^{13}C NMR (150 MHz, DMSO) 150.4 (C-2), 147.5, 146.4 (C-8?), 85.7 (C-1′), 85.4 (C-4′), 71.0 (C-3′), 70.3 (C-2′), 62.4 (C-5′), 40.4. LRMS (*m/z*): $[M+H^+]^+$ calcd for formula $C_{10}H_{14}N_5O_5^+$, 284.3; found 284.1. This data agrees with literature.[393]

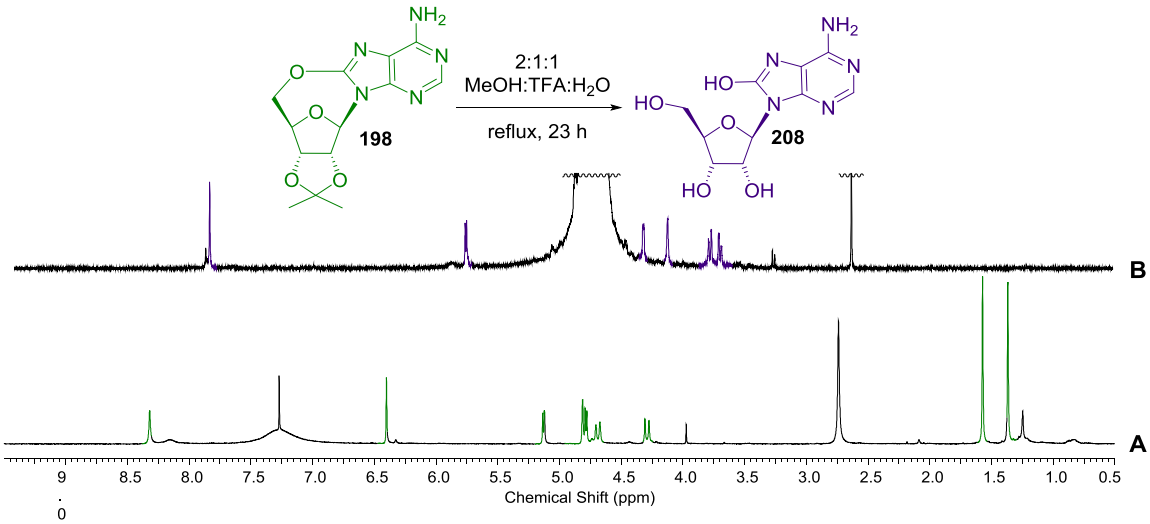


Figure 7.5.13: *^{1}H NMR spectra (600 MHz, 0.5-9.5 ppm) showing attempted deprotection of* ***198*** *(in green) in 2:1:1 MeOH:TFA:H_2O, after heating at reflux for:* ***A****) 5 h ($CDCl_3$); and* ***B****) a further 18 h (D_2O).*

	Capon[393] (500 MHz, DMSO)	Wilson[408] (500 MHz, DMSO)	Klostermeier[409] (300 MHz, D_2O)	Sasaki[410] (270 MHz, DMSO)	McCloskey[411] (? MHz, D_2O)	Meunier[412] (250 MHz, CD_3COOD)	Saladino[413] (? MHz, DMSO)
	10.39 (br s, 8-OH)	10.36 (s, 1H)		10.30 (br s, 1H)			10.34 (1H, br s, NH)
H-2	8.02 (s)*	8.01 (s, 1H)	8.04 (s, 1H)*	7.94 (1H, s)	8.04 (s)*	8.46 (1H, s)*	8.02 (1H, s)*
6-NH_2	6.55 (br s)*	6.56 (s, 2H)		6.49 (2H, br s)			6.59 (1H, br s, NH)
H-1′	5.68 (d, *J* = 6.1 Hz)*	5.66 (d, *J* – 6.5 Hz, 1H)	5.90 (d, J = 6.5 Hz, 1H)*	5.60 (1H, d, *J* = 2.0 Hz)	5.86 (d, *J* = 6.8 Hz)*	6.05 (1H, d, *J* = 4.5 Hz)*	5.67 (1H, d, *J* = 6.3 Hz)*
2′,3′,5′ OH	5.13 (br m)*	5.23 (d, *J* = 6.2), 5.17 (d, *J* = 4), 5.05 (d, *J* = 4.5)		5.17-5.19 (1H, m), 5.09-5.13 (1H, m), 4.99-5.00 (1H, m)			
H-2′	4.87 (br m)*	4.85 (dd, *J* = 11.7, 6.0)	4.99 (dd, J = 6.3, 5.6 Hz)*	4.79 – 4.82 (1H, m)	4.95 (dd, *J* = 6.8, 5.7 Hz)*	5.12 (1H, t, *J* = 4.5 Hz)*	4.86 (1H, m)*
H-3′	4.14 (br m)*	4.11 (d, *J* = 2.9)	4.48 (dd, J = 5.5, 3.2 Hz)*	4.05 (1H, br s)	4.39 (dd, *J* = 5.7, 2.8 Hz)*	4.84 (1H, br s)*	4.13 (1H, m)*
H-4′	3.87 (br m)*	3.85 (dd, *J* = 7, 4.0)	4.25 (dd, J = 6.6, 3.2 Hz)*	3.79 (1H, br s)	4.18 (m)*	4.3 (3H, br s, H-4′, H-5′a and b)	3.86 (1H, m)*
H-5′a	3.62 (br m)*	3.60 (bd, *J* = 12.0)	3.93 (dd, *J* = 12.7, 2.8 Hz)*	3.43 – 3.56 (2H, m)	3.84 (dd, *J* = 12.6, 2.8 Hz)*		3.60 – 2.94 (3H, m, H-5′a and b)
H-5′b	3.49 (br m)*	3.45 (ddd, *J* = 4.5, 7, 12)	3.84 (dd, J = 12.8, 3.9 Hz)*		3.75 (dd, *J* = 12.6, 3.6 Hz)*		

Table 7.5.7: Literature NMR data for 8-oxo-adenosine **208**

* denotes proton has been assigned by authors

Attempt with TFA in water:

General procedure 5.3: 2′,3′-*O,O*-Isopropylidene-8,5′-anhydro-8-oxyadenosine **198** was dissolved in water and TFA, stirred at rt for 5 h, and the solvent was removed *in vacuo.* The residue was subjected to ^{1}H NMR analysis (***Figure 7.5.14***).

Table 5.1 Entries 5 and 6: 2′,3′-*O,O*-Isopropylidene-8,5′-anhydro-8-oxyadenosine **198** (41.26 mg, 0.14 mmol) was dissolved in water (1 mL) and TFA (0.2 mL). The solution was divided equally into two flasks.

Table 5.1 Entry 5: further water (1 mL, bringing total to 1.5 mL) and TFA (0.4 mL, bringing total to 0.5 mL) were added. General procedure 5.3.

Table 5.1 Entry 6: further water (0.5 mL, bringing total to 1 mL) and TFA (0.9 mL, bringing total to 1 mL) were added. General procedure 5.3..

Table 5.1 Entry 7: 2′,3′-*O,O*-Isopropylidene-8,5′-anhydro-8-oxyadenosine **198** (16.12 mg, 0.05 mmol) in water (0.4 mL) and TFA (1.2 mL). General procedure 5.3.

Table 5.1 Entry 8: 2′,3′-*O,O*-Isopropylidene-8,5′-anhydro-8-oxyadenosine **198** (21.2 mg, 0.07 mmol) in TFA (1.8 mL) and water (0.2 mL). General procedure 5.3.

Entry 8 repeated with additional timepoints:

2′,3′-*O,O*-Isopropylidene-8,5′-anhydro-8-oxyadenosine **198** (21.4 mg, 0.07 mmol) was stirred in a mixture of TFA (1.8 mL), water (0.2 mL) for 1.5 h. The solvent was removed *in vacuo*. ^{1}H NMR analysis indicated that approximately 19% conversion to 8,5′-anhydro-8-oxyadenosine *ribo*-**146** had taken place. The sample was again stirred in a mixture of TFA (1.8 mL), water (0.2 mL) for 1.5 h, and the solvent removed *in vacuo*. ^{1}H NMR integral analysis indicated that approximately 50% conversion to *ribo*-**146** had taken place. The sample was again stirred in a mixture of TFA (1.8 mL), water (0.2 mL) for 2.5 h, and the solvent removed *in vacuo*. ^{1}H NMR integral analysis indicated that approximately 65% conversion to *ribo*-**146** had taken place (***Figure 7.5.15***).

^{1}H NMR (600 MHz, DMSO) δ 8.27 (1H, s, H-2), 5.99 (1H, s, H-1′), 4.64 (1H, dd, *J* = 12.9, 2.0 Hz, H-5′a) 4.58 (1H, s, H-4′), 4.45 (1H, d, *J* = 6.2 Hz, H-3′), 4.28 (1H, d, *J* = 6.2 Hz, H-2′), 4.14 (1H, d, *J* = 12.6 Hz, H-5′b). This data agrees with literature.[393,392]

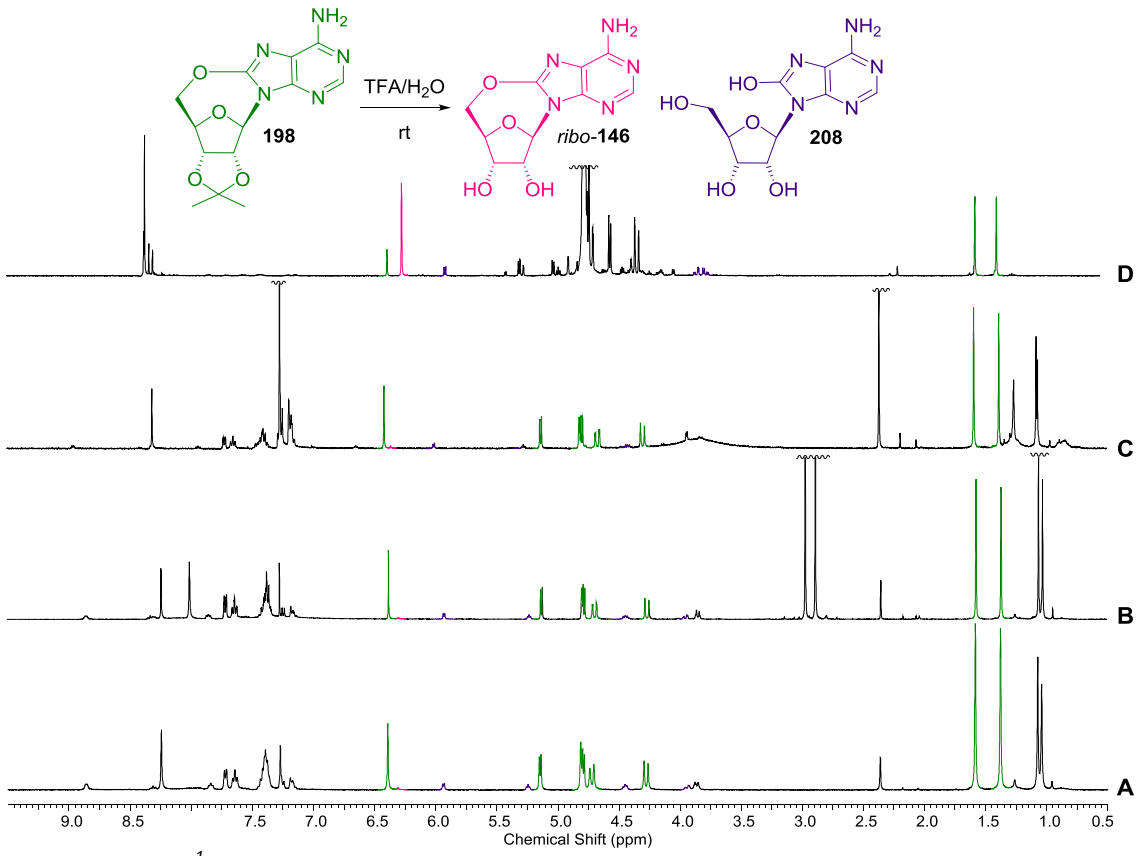


***Figure 7.5.14**: ^{1}H NMR spectra (400 MHz, 0.5-9.5 ppm) showing deprotections of **198** in TFA/H_2O after 5 h: **A**) 1:3 TFA:H_2O (**Table 5.1** entry 5, $CDCl_3$). ribo-**146:** δ 8.24 (overlaps with other species, H-2), 6.31 (s, H-1'). **208:** δ 8.24 (overlaps with other species, H-2), 5.94 (1H, d, J = 5.0, H-1'), 5.24 (1H, app t, J = 5.3 Hz), 4.44 (1H, dd, J = 6.8, 5.8 Hz), 3.96 – 3.91 (1H, m, H-5'a), 3.84 – 3.89 (1H, m, H-5'b, overlapping with unidentified species). **B**) 1:1 TFA:H_2O (**Table 5.1** entry 6, $CDCl_3$). ribo-**146:** δ 8.24 (overlaps with other species, H-2), 6.30 (s, H-1'). **208:** δ 8.24 (overlaps with other species, H-2), 5.92 (1H, d, J = 5.3, H-1'), 5.23 (1H, app t, J = 5.5 Hz), 4.44 (1H, dd, J = 7.8, 5.8 Hz), 3.96 – 3.912 (1H, m, H-5'a), 3.87 – 3.82 (1H, m, H-5'b, overlapping with unidentified species). **C**) 3:1 TFA:H_2O (**Table 5.1** entry 7, $CDCl_3$). ribo-**146:** δ 8.24 (overlaps with other species, H-2), 6.36 (s, H-1'). **208:** δ 8.24 (overlaps with other species, H-2), 6.01 (1H, d, J = 5.0, H-1'), 5.27 (1H, app t, J = 5.3 Hz), 4.42 (1H, dd, J = 8.5, 5.3 Hz). **D**) 9:1 TFA:H_2O (**Table 5.1** entry 8, D_2O). ribo-**146:** δ 6.25 (s, H-1'), 4.31 (1H, d, J = 13.3 Hz, H-5'b). **208:** δ 5.88 (1H, d, J = 5.5), 5.01 (1H, app t, J = 6.0 Hz, H-2'), 3.87 (1H, ABX, J = 12.6, 3.3 Hz, H-5'a), 3.80 (1H, ABX, J = 12.3, 5.3 Hz, H-5'b). Residual toluene (from co-evaporation during work-up) is visible in spectra A-C.*

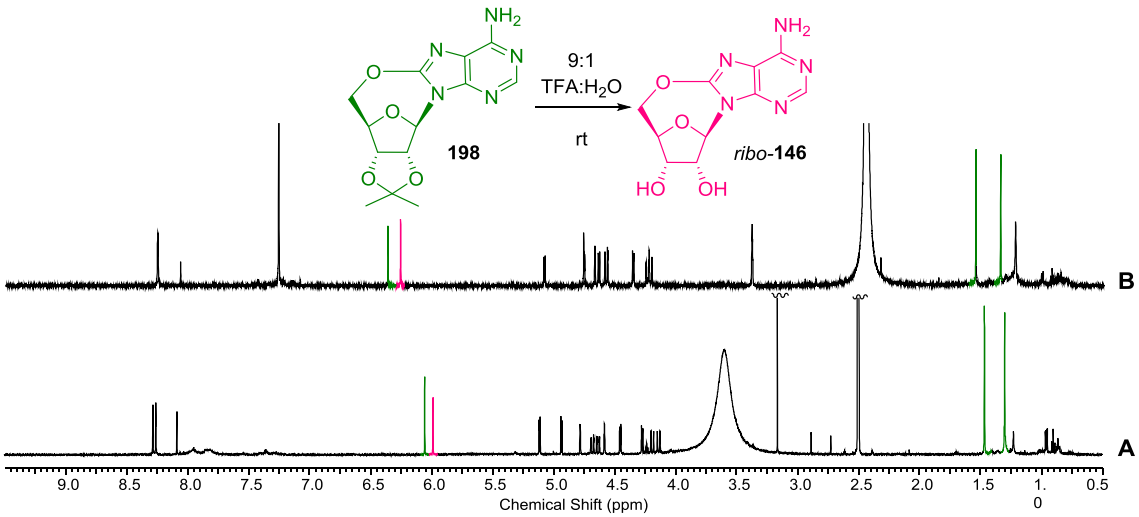
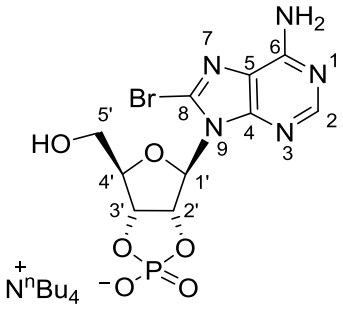


Figure 7.5.15: *Comparison of sections of ^{1}H NMR spectra (600 MHz, 0.5-9.5 ppm) of attempted deprotections of* ***198*** *with TFA.* ***A****) After 3h, in DMSO (****198*** *δ 6.05, s; ribo-****146*** *5.99, s); and* ***B****) after 5.5 h, suspended in $CDCl_3$ (****198*** *δ 6.39, s; ribo-****146*** *δ 6.27, s; the H-2 peaks have coalesced). A downfield shift in the H-1' peaks for both the starting material and product can be seen upon changing the NMR solvent from DMSO to $CDCl_3$.*

7.5.5 Synthesis of 8,5′-anhydro-8-oxy-2′,3′-cyclic-adenosine monophosphate *ribo-146-2',3'cP*

Tetra ^{n}butyl ammonium 8-bromo-2',3'-cyclic-adenosine monophosphate **214**-2′,3′cP



55-2′,3′cP (512 mg, 0.15 mmol) was stirred in a mixture of water (3.5 mL) and 1M NaOH (0.15 mL). 1M NaOAc buffer (pH 4, 5 mL) and then bromine water (1 mL, 0.27 mmol) were added and the resulting solution stirred in the dark for 4 h. Additional bromine water (1 mL, 0.32 mmol) was added and the solution stirred in the dark for a further 18 h. The bromination reaction was easily followed by ^{1}H NMR analysis (***Figure 7.5.16***). Saturated AgOAc aqueous solution was added to precipitate bromide and the resulting suspension filtered. The filtrate was acidified (pH 4.08 to pH 2.44) and washed with EtOAc (3 x 20 mL). The aqueous layer (pH 6.05) was lyophilised. The lyophilisate was dissolved in water (5 mL) and acidified and washed with EtOAc as before, and the aqueous layer concentrated by rotary evaporation (^{1}H NMR analysis

indicated around 20 eq. acetate remained). The sample was purified by flash column chromatography (^{i}PrOH:H_2O 90:10 to 75:25), the fractions containing **213-**2′,3′cP combined and the solvent removed *in vacuo* and by lyophilisation. Another purification by automated flash column chromatography (reverse-phase C-18, MeOH: H_2O 2.5:97.5 to 12.5:87.5) was carried out, the fractions containing **213-**2′,3′cP combined and the solvent removed *in vacuo* and by lyophilisation (this step resulted in slight cleavage of the cyclic phosphate, shown in ***Figure 7.5.17***). The sample was then dissolved in water and stirred with $N^{n}Bu_4$-dowex for 2 h, filtered, and lyophilised. The sample was purified again by automated flash column chromatography (reverse-phase C-18, MeOH: H_2O 2.5:97.5 to 25:75), the fractions containing the desired product combined and the solvent removed *in vacuo* and by lyophilisation to provide tetra ^{n}butyl ammonium 8-bromo-2′,3′-cyclic adenosine monophosphate **214**-2′,3′cP (4.2 mg, 4%) as a white powder.

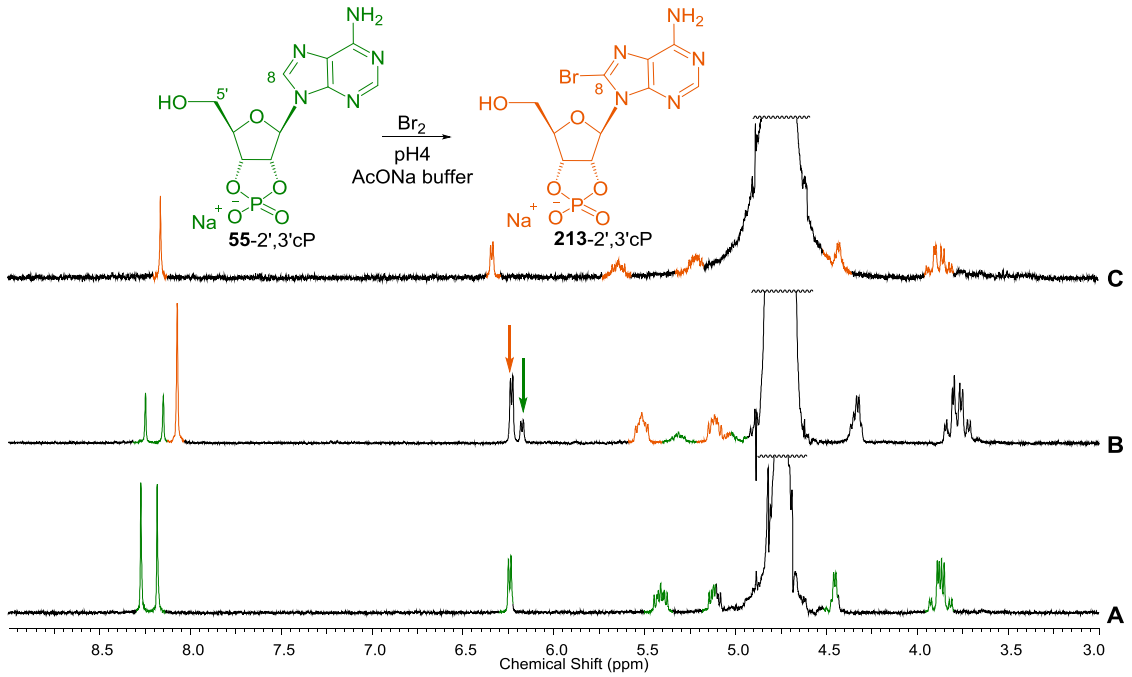


Figure 7.5.16: *^{1}H NMR spectra (300 MHz, D_2O, 3.0-9.0 ppm): the disappearance of 8-H, and the respective upfield and downfield shifts of the 2-H and 1'-H meant that the bromination of **55**-2',3'cP could be easily followed by ^{1}H NMR. **A**) Commercial **55**-2',3'cP; **B**) 75% conversion can be observed after 23 h (0.34 mmol scale, 7.2 eq Br); **C**) completed reaction after work-up. Slight differences in chemical shift are likely due to differences in pH.*

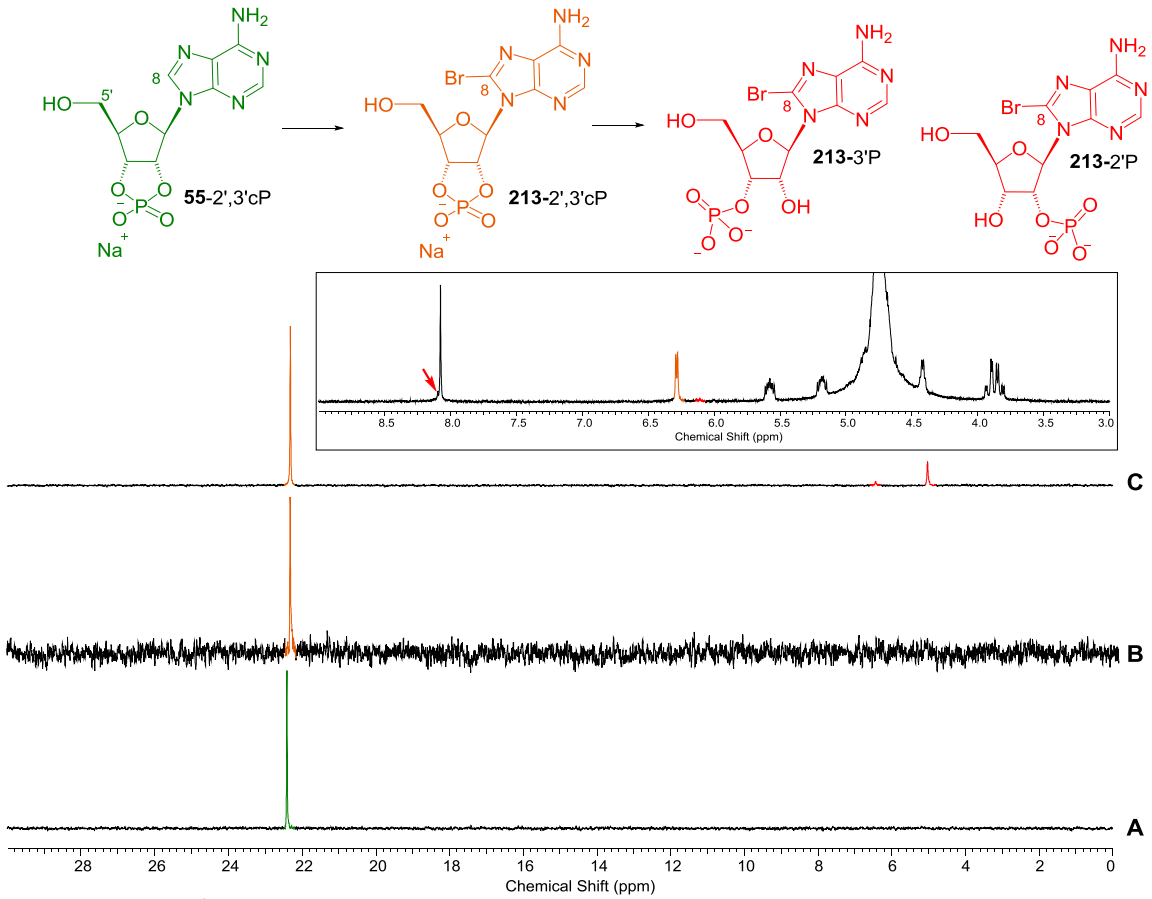


***Figure 7.5.17**: ^{31}P NMR spectra (121 MHz, D_2O, 0.0-30.0 ppm) showing formation and unintended partial degradation of **214**-2',3'cP. **A**) starting material **55**-2',3'cP, δ 20.46 ; **B**) **214**-2',3'cP after first purification by flash column chromatography, δ 22.36; **C**) after first purification by reverse-phase flash column chromatography. **213**-2',3'cP (δ 22.32, 76 % by relative integration phosphorylated species) has undergone partial conversion to the ring-opened isomers, characterised by peaks at δ 6.43 (4%) and δ 5.02 (20%). Inset: ^{1}H NMR spectrum (300 MHz, 3.0-9.0 ppm) after first purification by reverse-phase flash column chromatography. Only two anomeric protons are apparent: δ 6.12 (d, J = 7 Hz, 5% by relative integration) and δ 6.28 (d, J = 4 Hz, 95% by relative integration). It is likely that this latter peak obscures the anomeric proton of the major ring-opened product.*

*Synthesis from sodium 2',3'-cyclic adenosine monophosphate **55**-2',3'cP*, 2nd attempt:

55-2',3'cP (117.7 mg, 0.34 mmol) was stirred in a mixture of water (7 mL) and 1M NaOH (0.3 mL). 1M NaOAc buffer (pH 4, 10 mL) and then bromine water (2 mL, 1.75 mmol) were added and the resulting solution stirred in the dark for 23 h. Additional bromine water (2 mL, 0.71 mmol) was added and the solution stirred in the dark for a further 22 h. Saturated AgOAc aqueous solution was added to precipitate bromide and the resulting suspension filtered. The filtrate was acidified (pH 3.88 to pH 2.44) and

washed with EtOAc (3 x 50 mL). The aqueous layer (pH 3.99) was concentrated *in vacuo* to half its volume and then lyophilised. The sample was then dissolved in water and stirred with N^nBu_4-dowex for 2 h, filtered, and lyophilised. The sample was then purified by automated flash column chromatography (reverse-phase C-18, MeOH: H_2O 2.5:97.5 to 25:75), but NMR analysis of all fractions revealed complete hydrolysis of the cyclic phosphate had taken place.

*Synthesis from 2′/3′- adenosine monophosphate mixed isomers **55**-2′P and **55**-3′P*

55-2′P and **55**-3′P (225.9 mg, 0.65 mmol) was stirred in a mixture of 1M NaOAc (20 mL) and 1M NaOH (0.6 mL). Bromine water (18.4 mL, 7.8 mmol) was added in the resulting solution stirred in the dark for 48 h. Saturated AgOAc aqueous solution was added to precipitate bromide and the resulting suspension filtered. The filtrate was acidified (pH 3.76 to pH 2.04) and washed with EtOAc (3 x 40 mL). The aqueous layer was lyophilised. The lyophilisate was dissolved in water (10 mL) and stirred with N^nBu_4-dowex overnight, filtered, and lyophilised. The sample was dissolved in water (25 mL) and EDC (391.1 mg, 2.04 mmol) added. The addition of EDC raised the pH from 2.58 to 4.19. The resulting solution was stirred for 2 h and then lyophilised. The sample was then purified by automated flash column chromatography (reverse-phase C-18, MeOH: H_2O 2.5:97.5 to 25:75), the fractions containing the desired product combined and the solvent removed *in vacuo* and by lyophilisation to provide **55** (50.6 mg, 35%) as a white powder. ^{1}H NMR (600 MHz, D_2O) δ 8.22 (1H, s, H-2), 6.39 (1H, d, *J* = 3.7 Hz, H-1′), 5.67-5.72 (1H, m, H-2′) 5.25-5.30 (1H, m, H-3′) 4.48 (1H, d, *J* = 3.9 Hz), 3.97 (1H, dd, *J* = 13.9, 2.8 Hz, H-5′a), 3.89 (1H, dd, *J* = 10.6, 4.6 Hz, H-5′a), 3.20-3.23 (8H, m, N(CH_2 CH_2 CH_2 CH_3)$_4$), 1.64-1.70 (8H, m, N(CH_2 CH_2 CH_2 CH_3)$_4$), 1.35-1.40 (8H, m, (CH_2 CH_2 CH_2 CH_3)$_4$), 0.95-0.98 (12H, m, N(CH_2 CH_2 CH_2 CH_3)$_4$). ^{13}C NMR (150 MHz, D_2O) δ 153.99 (C-2). 128.42, 91.14 (C-1′), 86.01 (C-4′), 80.22 (C-2′), 78.19 (C-3′), 61.82 (C-5′), 58.75 (N(CH_2 CH_2 CH_2 CH_3)$_4$), 23.77(N(CH_2 CH_2 CH_2 CH_3)$_4$), 19.81 (N(CH_2 CH_2 CH_2 CH_3)$_4$), 13.47(N(CH_2 CH_2 CH_2 CH_3)$_4$). ^{31}P NMR (121 MHz, D_2O) δ 22.35 (C-2′/3′). M.p. >165 °C (decomposes). IR(cm^{-1}) 3030-3050 (O-H, N-H), 2875-2961 (C-H), 1586 (N-H). LRMS (*m/z*): $[M+H^+]^+$ calcd for formula $C_{10}H_{10}N_5O_5PBr^-$, 406.19; found 405.86. These nucleotide ^{13}C chemical shifts agree with literature.[414]

*Unsuccessful formation of sodium 2',3'-cyclic adenosine monophosphate **55**-2',3'cP from 2'/3'- adenosine monophosphate (mixed isomers) **55**-2'P and **55**-3'P:*

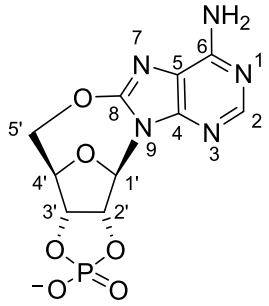
Attempt 1 (forming the cyclic phosphate first, followed by bromination and ion exchange of the cyclic phosphate): **55**-2′P and **55**-3′P (235.9 mg, 0.68 mmol) was dissolved in water (20 mL) and the pH adjusted from 2.89 to 3.66. EDCI (183.7.0 mg, 0.96 mmol) was dissolved in the solution and the pH further adjusted to 3.87. The reaction mixture was adjusted from pH 3.4 to 4.0 after 3.5 h, when NMR analysis indicated 85% converison. Additional EDCI (22.3 mg, 0.12 mmol) was added after 5.5 h (when NMR analysis indicated 90% conversion), and the solution adjusted to pH 4.3. The solution stirred for a further 18.5 h, then lyophilised. NMR analysis revealed that **55**-2′P and **55**-3′P were present in 25% yield by relative integration. Purification by automated flash column chromatography (reverse-phase C-18, MeOH: H_2O 2.5:97.5 to 25:75) was carried out, but did not succeed in separating the desired product.

Attempt 2 (cyclic phosphate formed, then attempted purification):
55-2′P and **55**-3′P (248.0 mg, 0.68 mmol) was dissolved in water (20 mL) and the pH adjusted from 2.60 to 3.81. EDCI (222.0 mg, 1.15 mmol) was dissolved in the solution and the pH further adjusted to 4.14. NMRs were stirred periodically acquired, and indicated the reaction was complete after 3.5 h. The solution was lyophilised and purification by automated flash column chromatography (reverse-phase C-18, MeOH: H_2O 2.5:97.5 to 25:75) was carried out, the fractions containing **55**-2′,3′cP combined and the solvent removed in vacuo and by lyophilisation. NMR analysis revealed the column had not succeeded in separating EDCI from the nucleotides. The resulting white powder was dissolved in a minimum quantity of water, loaded onto a column of proton form dowex (it was hoped that the positively charged amine in EDCI would ion exchange with proton and stick to negatively charged dowex). The column was eluted with 30 mL fraction of HCl (0 to 2 M). Each fraction was concentrated in vacuo to 10 mL, diluted with water (20 mL) and re-concentrated (x5) before being diluted to 30 mL and lyophilised. NMR analysis revealed the column had succeeded in separating EDCI from the nucleotides, but had also resulted in complete hydrolysis of the cyclic phosphate.

Attempt 3 (cyclic phosphate formed, then ion exchange):
55-2′P and **55**-3′P (310.2 mg, 0.85 mmol) and EDCI (380.0 mg, 1.98 mmol) were stirred in water (5 mL) for 2 h. The mixture was then stirred with an excess of N^nBu_4-

dowex for 30 mins, then filtered, rinsed with water (2 mL), and the combined aqueous filtrate and washings lyophilised. NMR analysis revealed that partial hydrolysis (48%) of the cyclic phosphate had taken place.

*Attempted formation of 8,5′-anhydro-8-oxy-2′,3′-cyclic-adenosine monophosphate ribo-**146**-2′,3′cP*[381,367]



Tetra ^{n}butyl ammonium 8-bromo-2′,3′-cyclic-adenosine monophosphate **214**-2′,3′cP (10.1 mg, 0.02 mmol) was stirred in ^{t}BuOH (1 mL) and ^{t}BuOK (5.2 mg, 0.05 mmol). NaH (17.7 mg, 60 % in mineral oil, 0.44 mmol) was washed with dry hexane and added to the suspension. The suspension was stirred for 3.5 h, then EtOH (0.5 mL) and H_2O (1 mL) added. Solvent was removed *in vacuo* at room temperature and by lyophilisation. ^{1}H NMR analysis showed that *ribo*-**146**-2′,3′cP had not formed; ^{31}P NMR analysis revealed that the cyclic phosphate had opened (***Figure 7.5.18***).

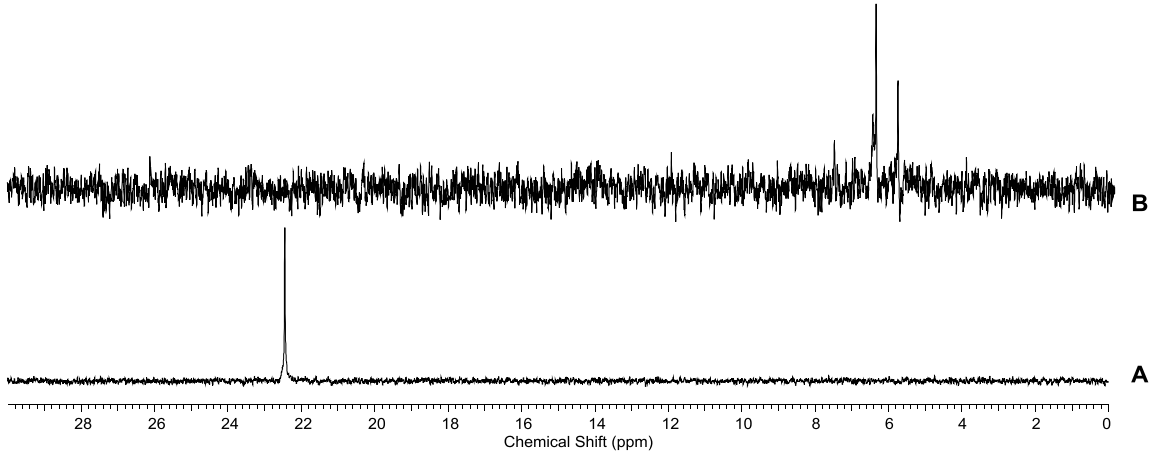


***Figure 7.5.18**: ^{31}P NMR spectra (121 MHz, D_2O, 0.0-30.0 ppm) showing unintended cyclic phosphate hydrolysis during attempted 5′,8-anhydro bond formation.*

7.5.6 Attempted synthesis of 8,5′-anhydro-8-oxy-guanosines

*Attempted formation of 8,5'-anhydro-8-oxy-guanosine with DNPA **215***

Attempt 1: Guanosine ß-**56** (30.6 mg, 0.11 mmol) and DNPA **215** (81.9 mg, 0.41 mmol) were stirred in a mixture of D_2O (250 µL) and DMF (750 µL) at 45 °C for 17 h. The solvent was removed *in vacuo*, and the residue dissolved in 0.1 M HCl (2 mL), washed with EtOAc (3 x 3 mL) and lyophilised. The lyophilsate was resuspended in D_2O and subjected to NMR analysis, which indicated a complex mixture of products were present (***Figure 7.5.19***).

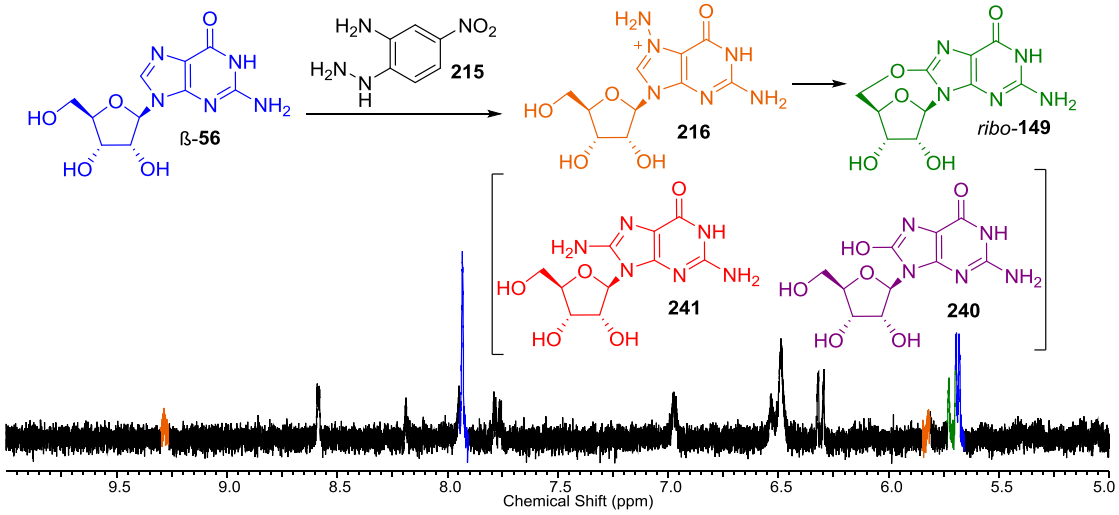


Figure 7.5.19: *^{1}H NMR spectrum (600 MHz, D_2O, 5.0-10.0 ppm) showing reaction of DNPA **215** with guanosine (**56**, blue) initially leads to the synthesis of N-7-aminoguanosine (**216**, orange), which has a distinctively deshielded C-8 proton (attributed here to the peak seen at δ 9.28, and reported at δ 9.22 by Kohda[385] and δ 9.08 by Kadlubar)[386]. A broad singlet at δ 5.82 has been attributed to the anomeric proton of this species: this proton has been variously reported as δ 5.82 (s)[385] and δ 5.73 (d, but no coupling constant provided).[386] A singlet at δ 5.73 (green) could be attributed to anomeric proton of the desired product, 5',8-anhydro guanosine ribo-**149**: Kohda reports this proton as a singlet at δ 5.72, although Kadlubar reports it as a doublet at δ 5.75 (coupling constant not provided). This peak, however, could also be the anomeric proton of other reported products: 8-oxo-guanosine **240** (expected at δ 5.58) or C-8-amino-guanosine **241** (reported at δ 5.73; both reported as doublets but without coupling constants).[386] It is not clear what species the other peaks observed represent (protons of the amino groups in all species are expected in the range δ 5.9-6.9,[385,386] but aromatic protons are only expected for species without C-8 substituents), indicating that a complex mixture of other by-products has formed.*

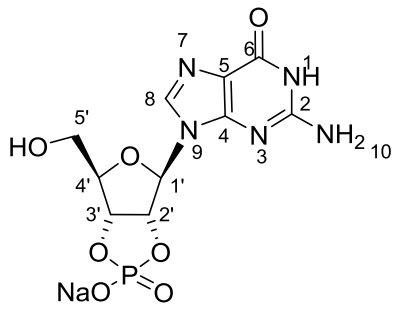
Attempt 2: Guanosine ß-**56** (118.9 mg, 0.42 mmol) and DNPA **215** (321.6 mg, 1.62 mmol) were stirred in a mixture of D_2O (2 mL) and DMF (6 mL) at 45 °C for 13 d. The

solvent was removed *in vacuo*, and the residue dissolved in 0.1 M HCl (8 mL), washed with EtOAc (2 x 8 mL) and lyophilised. The lyophilsate was resuspended in D_2O and subjected to NMR analysis, which indicated a complex mixture of products were present.

Tetraplatin **217**[415]

K_2PtCl_6 (62 mg, 0.13 mmol) was dissolved in 1M HCl (29 mL) and evaporated to near-dryness in vacuo. A solution of trans-1,2-diaminocyclohexane (30 mg, 0.26 mmol) in water (20 mL) was added. The solution was protected from light and stirred at reflux for 16h. The solution was cooled and further amine solution was added (34 mg, 0.3 mmol in 1 mL water). The solution was refluxed for a further 4 h. The water was evaporated in vacuo and the resultant solid dissolved in warm acetone (30 °C, 10 x 20 mL) and filtered through a cotton wool plug. The acetone solution was concentrated to 2 mL in vacuo and diluted with diethyl ether (20 mL). The resulting suspension was centrifuged and the supernatant pipetted off. Addition of diethyl ether, centrifugation and removal of the supernatant was repeated 3 times, and the solid yellow residue vacuum dried to afford tetraplatin **217** (35 mg, 0.08 mmol, 59%). LRMS (CI) 483.2 ($[M+2(NH_3)]^+$)

*2',3'-cyclic guanosine monophosphate **56**-2',3'cP*



2′/3′-guanosine monophosphate mixed isomers **56**-2′P and **56**-3′P (227 mg, 0.56 mmol) were dissolved in water (20 mL) and the pH lowered from pH 6.9 to pH 3.2 with 1M HCl. The addition of EDC (186 mg, 0.97 mmol) raised the pH to 3.8. The solution was stirred for 6 h at room temperature, when the addition of further EDC (20.0 mg, 0.10 mmol) raised the pH from pH 4.0 to 4.2. The solution was stirred for a further 19 h, then stirred with Na^+-dowex for 2 h, filtered, and lyophilised to give **56**-2′,3′cP as a white powder. ^{1}H NMR (600 MHz, DMSO) δ 11.00 (1H, br s, OH or NH), 7.95 (1H, s, H-8), 6.89 (2H, br s, 2-NH_2), 5.92 (1H, d, *J* = 4.4 Hz, H-1′), 5.01-4.79 (1H, m, H-2′), 4.68 (1H, ddd, *J* = 9.1, 7.0, 4.4 Hz, H-3′), 4.09 (1H, app q, *J* = 4.5 Hz, H-4′), 3.60 (1H, dd, *J* = 11.9, 4.0 Hz, H-5′a), 3.60 (1H, dd, *J* = 12.1, 5.1 Hz, H-5′b). ^{13}C NMR (150 MHz, DMSO)

δ 158.34, 156.73, 154.13, 151.07, 135.66 (C-8), 116.52, 87.47 (C-1′), 85.09 (C-4′), 78.69 (C-2′), 75.64 (C-3′), 61.36 (C-5′).). M.p. > 166 °C (decomposes). IR (cm^{-1}) 3500-3000 (O-H, N-H), 1694 (N-H). HRMS (*m/z*): $[M+H^+]^+$ calcd for formula $C_{10}H_{13}N_5O_7P$, 346.0553; found 346.0555.

Reactions of tetraplatin and guanosine-derivatives.

Solutions of **217** were prepared by dissolving **217** in D_2O in a foil-covered Eppendorf tube. Several heating (maximum 50 °C) and vortexing cycles over 2 or 3 h were required for total dissolution. Reaction solutions were filtered through a syringe-end filter disk (Kinesis KX Syringe Filter, PVDF, 0.45 µm pore size) before NMR analysis, and, except where specified, pH analysis.

*Reaction of **217** and 3′-deoxyguanosine monophosphate **218**-3′P*

An 11.1 mM solution of **218**-3′P in D_2O (1 mL) was added to a 10.6 mM solution of tetraplatin in D_2O (1 mL) and the pH adjusted to 8.6. The solution was stirred for 6 h at 37 °C. The appearance of a new species in the ^{1}H NMR spectrum (***Figure 7.5.20,*** H-1′: δ 6.57, d, *J* = 6.1 Hz) corresponds to the observations of Choi *et al*. (***Table 7.5.7***), indicating that 3′-deoxyguanosine monophosphate **218**-3′P was partially converted to 8,5′-anhydro-8-oxo-3′-deoxyguanosine monophosphate **220**.

Choi's assignments	Choi's data	notes	Observed resonances
220, H-1′	δ 6.47 (d)[a]	Appear simultaneously	δ 6.43 (d, *J* = 6.1 Hz)
$[Pt^{II}Cl_2(dach)]$ **223**	δ 2.24		-[b]
$[Pt^{IV}Cl_4(dach)]$ **217**	δ 2.4	Decreases over time	-[b]
224, H-8	δ 9.11		δ 9.08-9.12 (m)
225, H-8	δ 8.49		δ 8.40
226, H-8	δ 8.45	Appear at same time (from reaction of **223** with free **218**-3′P) with concomitant decrease in size of the peaks attributed to **218**-3′P	δ 8.33
226, H-1′	δ 6.26		-[b]
218-3′P, H-8	δ 8.17		δ 7.96
218-3′P, H-1′	δ 6.32		δ 6.27 (br s)

Table 7.5.7: Comparison of literature and observed ^{1}H NMR resonances in the reaction of 66 with 217

[a] coupling constants not reported.[384]

[b] could not be distinguished from other peaks in the same region

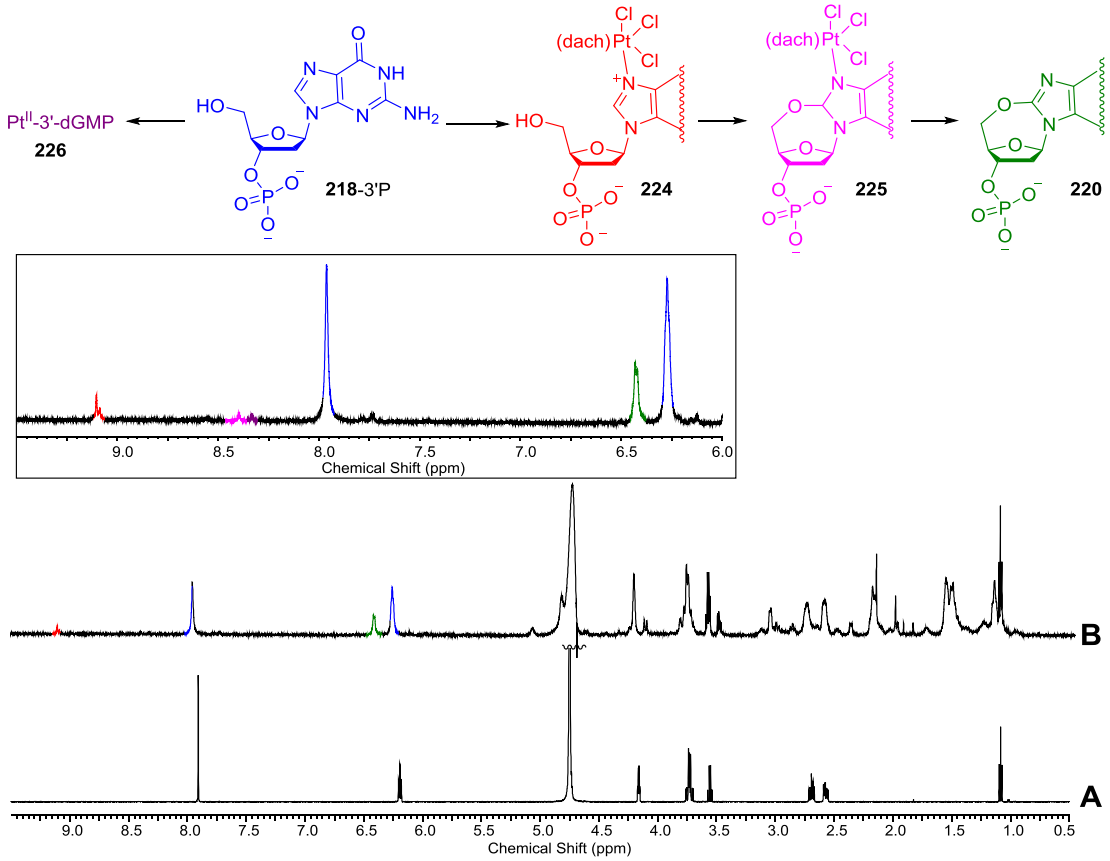


***Figure 7.5.20**: Reaction of deoxyguanosine 3'-monophosphate **218**-3'P with **217**. **Top**: **66** reacts with **217** to give the Pt^{IV}-N7 intermediate (**224**, red). Attack of the 5'-OH species at C-8 results in the formation of the Pt^{II}(5',8-anhydro) intermediate (**225**, pink). Loss of $[Pt^{II}Cl_2(dach)]$ **223** releases 8,5'-anhydro-8-oxo-3'-deoxyguanosine monophosphate (**220,** green); $[Pt^{II}Cl_2(dach)]$ can then react with free **218**-3'P to provide **226**, purple. **Bottom:** ^{1}H NMR spectra (600 MHz, D_2O, 0.5-9.5 ppm) showing the reaction of deoxyguanosine 3'-monophosphate **218**-3'P with **217**. **A**) Commercial deoxyguanosine 3'-monophosphate; and **B**) after reaction with **217** for 6 h. Inset: 6.0-9.5 ppm.*

*Reaction of **217** and 2′/3′-guanosine monophosphate mixed isomers **56**-2′P and **56**-3′P*

A 19.2 mM solution of **56**-2′P and **56**-3′P in D_2O (0.5 mL) was added to a 10.4 mM solution of **217** in D_2O (1 mL) and the pH adjusted to 8.7, and the solution stirred at 37 °C. At various timepoints, aliquots of the reaction mixture were drawn off for pH and ^{1}H NMR analysis (***Figure 7.5.21***).

*Reaction of **217** and guanosine **7***

An 18.7 mM solution of **7** in D_2O (0.3 mL) was added to a 9.5 mM solution of **217** in D_2O (1 mL), adjusted to pH 8.8, and stirred at 37 °C. At various timepoints, aliquots of the reaction mixture were drawn off for pH and ^{1}H NMR analysis (***Figure 7.5.22***).

*Reaction of **217** and 2′,3′-cyclic guanosine monophosphate **56**-2′,3′cP*

A 20.2 mM solution of **56**-2′,3′cP in D_2O (0.5 mL) was added to a 9.0 mM solution of **217** in D_2O (1.25 mL), adjusted to pH 8.6, and stirred at 37 °C. At various timepoints, aliquots of the reaction mixture were drawn off for pH and ^{1}H NMR analysis (***Figure 7.5.23***).

*Reaction of **217** and 2′,3′-cyclic guanosine monophosphate **56**-2′,3′cP with pH manually controlled*

A 19.0 mM solution of **56**-2′,3′cP in D_2O (0.5 mL) was added to a 13.8 mM solution of **217** in D_2O (0.75 mL), adjusted to pH 8.5, and stirred at 37 °C. At various timepoints, the pH of the reaction mixture was measured and adjusted without filtration, and aliquots removed and filtered for ^{1}H NMR analysis (***Figure 7.5.25***). After stirring for 40 mins, the pH of the reaction mixture was adjusted from 7.6 to 8.7; after 2 h the pH was adjusted from 7.9 to 8.3, and after 4 h from 7.9 to 8.5.

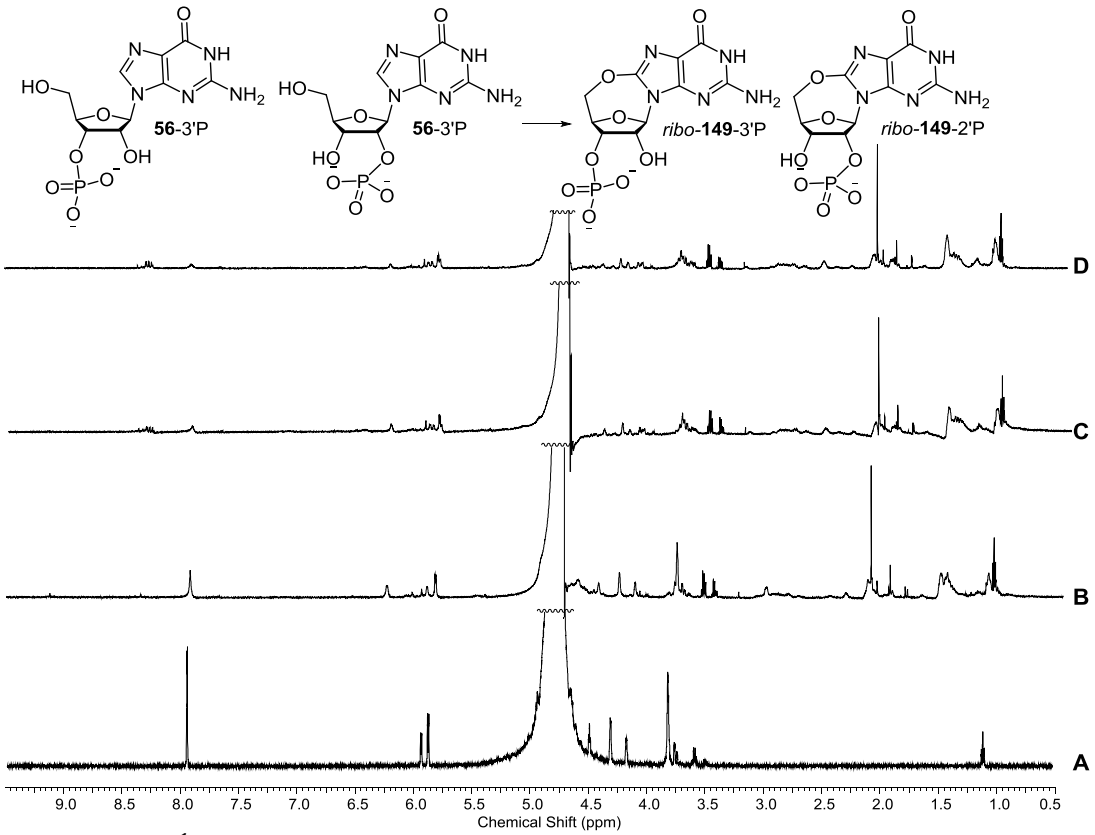


***Figure 7.5.21**: ^{1}H NMR spectra (400 MHz, D_2O, 0.5-9.5 ppm) showing the reaction of **56**-2'P and **56**-3'P with **217** at various time-points. **A**) Commercial **56**-2'P and **56**-3'P; and after reacting with **217** at 37 °C for: **B**) 4 h (pH 5.6); **C**) 21 h (pH 5.2); and **D**) 26 h, pH 5.2. The complexity of the reaction mixture increased as time progressed, but the identity of the products are unknown.*

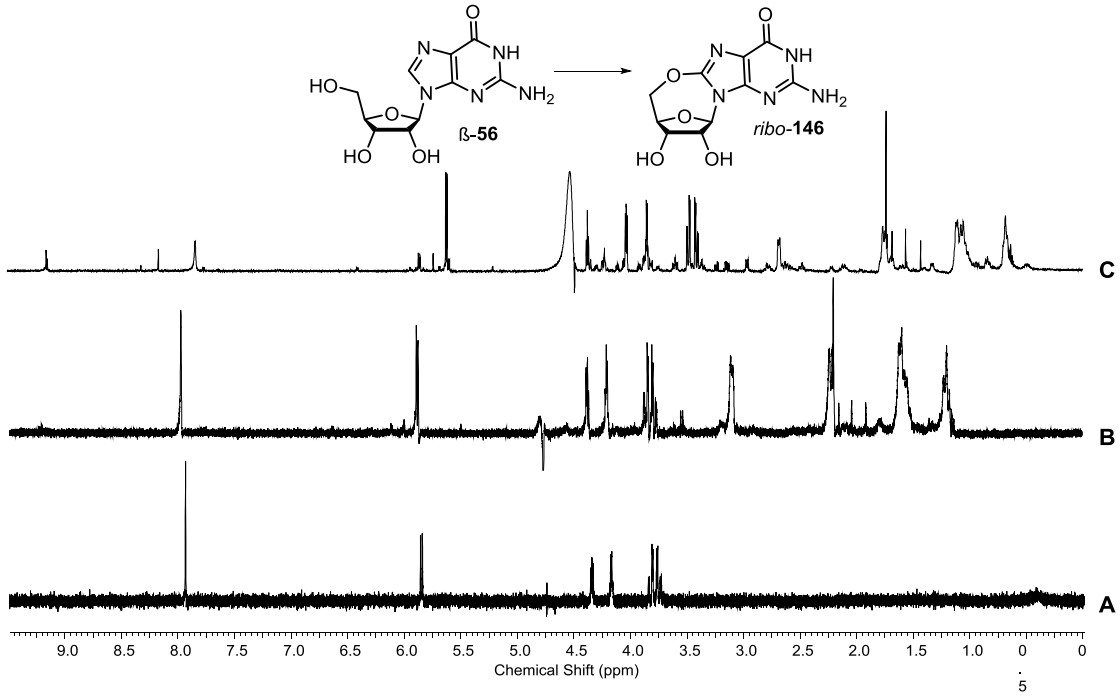


***Figure 7.5.22**: [1]H NMR spectra (400 MHz, D_2O, 0.0-9.5 ppm) showing the reaction of ß-**56** with **217** at various time-points. **A**) Commercial ß-**56**; and after reacting with **217** at 37 °C for: **B**) 4 h (pH 4.0); and **C**) 22 h (pH 3.9). The 1'-H region of the reaction mixture appears to be very complex.*

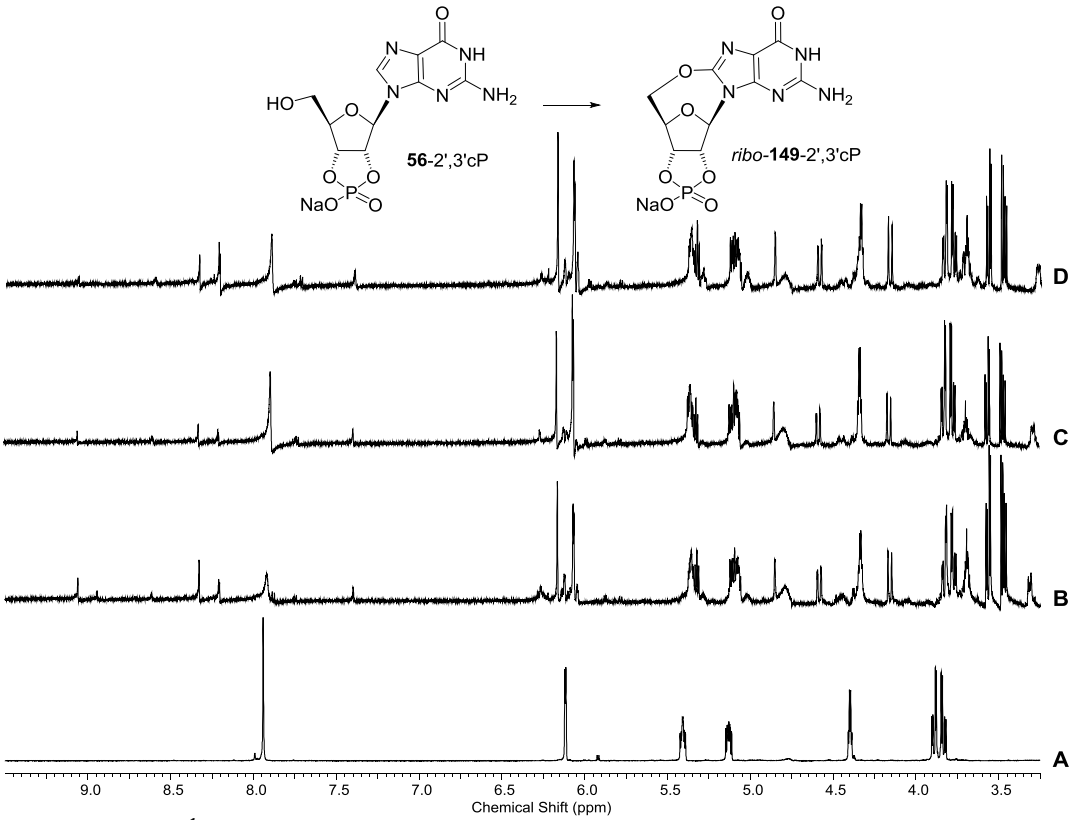


***Figure 7.5.23**: ^{1}H NMR spectra (400 MHz, D_2O, 3.25-9.5 ppm) showing the reaction of **56**-2',3'cP with **217** at various time-points. **A**) **56**-2',3'cP; and after reacting with **217** at 37 °C for: **B**) 2 h (pH 4.1); **C**) 3 h (pH 4.3); and **D**) 23 h (pH 4.3).*

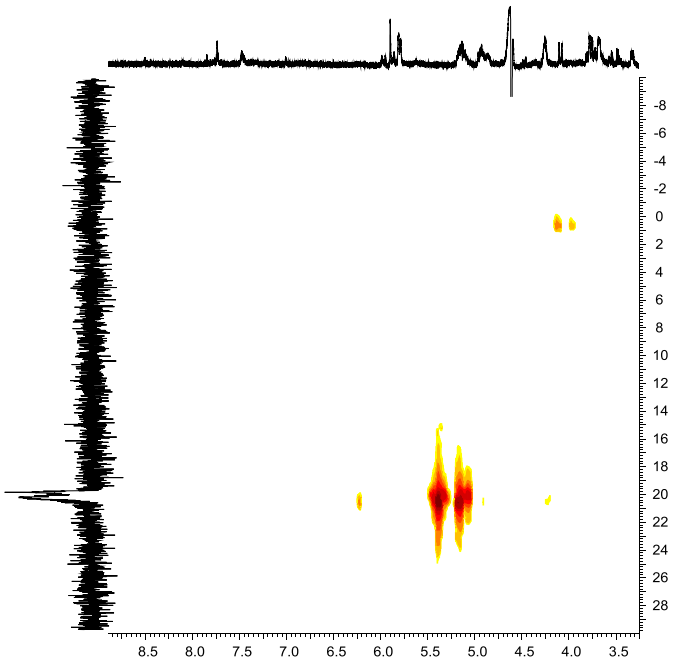


***Figure 7.5.24:** ^{31}P HMBC NMR spectrum (400 MHz, D_2O) showing the presence of a cyclic phosphate after the reaction of **56**-2',3'cP with **217** after 23 h. The increased complexity seen in the ^{1}H NMR spectrum is not due to hydrolysis of the cyclic phosphate.*

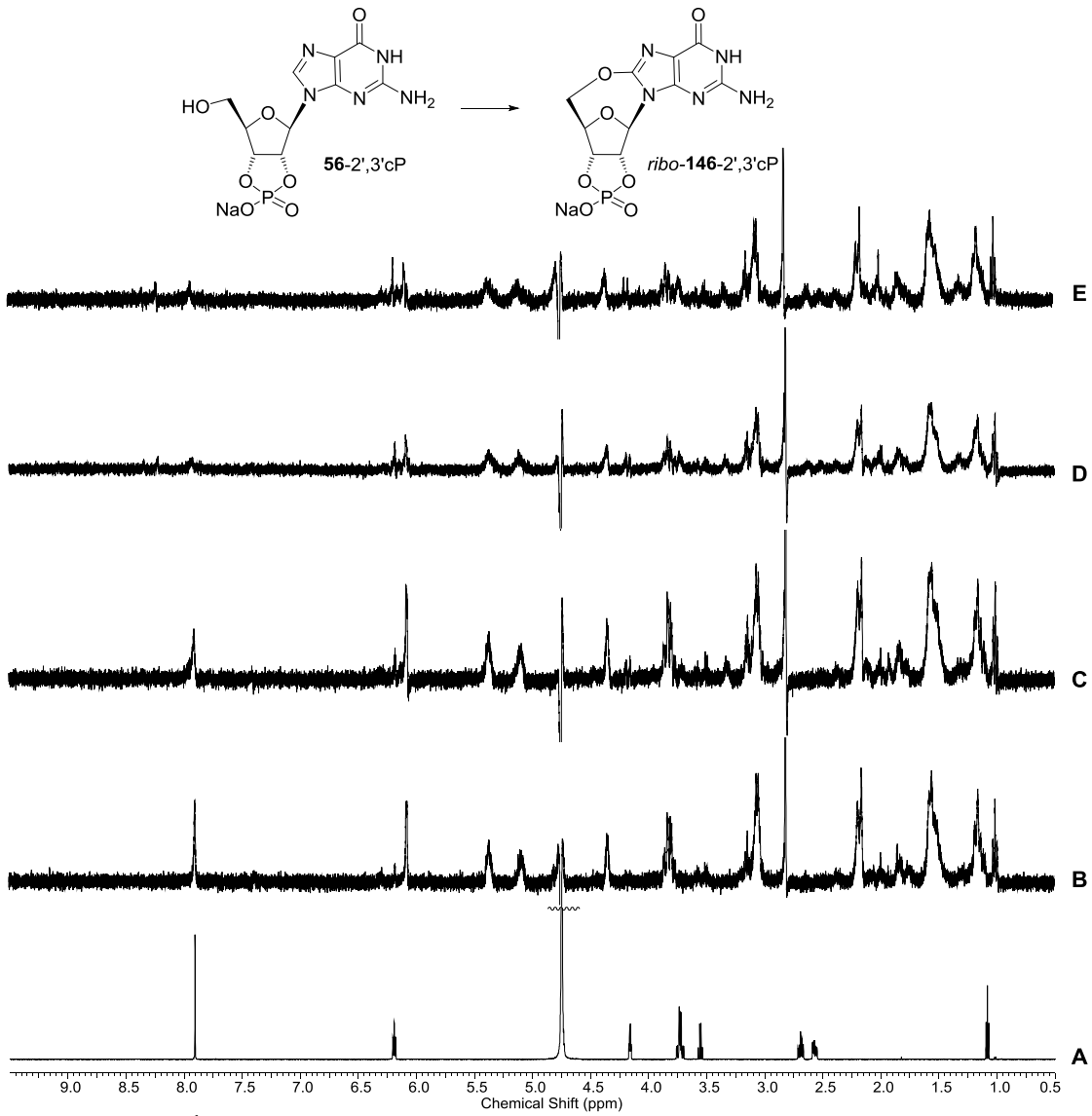


***Figure 7.5.25**: ^{1}H NMR spectra (400 MHz, D_2O, 0.5-9.5 ppm) showing the pH-controlled reaction of **56**-2',3'cP with **217** at various time-points. **A**) **56**-2',3'cP; and after reacting with **217** at 37 °C for: **B**) 40 min (pH 7.8 to 8.7); **C**) 2 h (pH 7.9 to 8.3); **D**) 4 h (pH 7.9 to 8.5); and **E**) 6 h.*

8. References

1. Deamer, D. & Weber, A. L. Bioenergetics and Life's Origins. *Cold Spring Harb. Perspect. Biol.* **2**, a004929 (2010).
2. Sutherland, J. D. The Origin of Life—Out of the Blue. *Angew. Chem. Int. Ed.* **55**, 104–121 (2016).
3. Vaneechoutte, M. & Fani, R. From the primordial soup to the latest universal common ancestor. *Res. Microbiol.* **160**, 437–440 (2009).
4. Penny, D. An Interpretive Review of the Origin of Life Research. *Biol. Philos.* **20**, 633–671 (2005).
5. Orgel, L. E. The origin of life—a review of facts and speculations. *Trends Biochem. Sci.* **23**, 491–495 (1998).
6. Sarafian, A. R., Nielsen, S. G., Marschall, H. R., McCubbin, F. M. & Monteleone, B. D. Early accretion of water in the inner solar system from a carbonaceous chondrite–like source. *Science* **346**, 623–626 (2014).
7. Kasting, J. F. Earth's early atmosphere. *Science* **259**, 920–926 (1993).
8. Thaddeus, P. The prebiotic molecules observed in the interstellar gas. *Philos. Trans. R. Soc. B Biol. Sci.* **361**, 1681–1687 (2006).
9. Sagan, C. & Chyba, C. The Early Faint Sun Paradox: Organic Shielding of Ultraviolet-Labile Greenhouse Gases. *Science* **276**, 1217–1221 (1997).
10. Crowe, S. A. *et al.* Atmospheric oxygenation three billion years ago. *Nature* **501**, 535–538 (2013).
11. Ranjan, S. & Sasselov, D. D. Constraints on the Early Terrestrial Surface UV Environment Relevant to Prebiotic Chemistry. *ArXiv161006223 Astro-Ph* (2016).
12. Ranjan, S. & Sasselov, D. D. Influence of the UV Environment on the Synthesis of Prebiotic Molecules. *Astrobiology* **16**, 68–88 (2016).

13. Leslie E. Orgel, O. Prebiotic Chemistry and the Origin of the RNA World. *Crit. Rev. Biochem. Mol. Biol.* **39**, 99–123 (2004).
14. Schopf, J. W. Fossil evidence of Archaean life. *Philos. Trans. R. Soc. B Biol. Sci.* **361**, 869–885 (2006).
15. Buick, R. Carbonaceous filaments from North Pole, Western Australia: Are they fossil bacteria in Archaean stromatolites? *Precambrian Res.* **24**, 157–172 (1984).
16. Wächtershäuser, G. Towards a Reconstruction of Ancestral Genomes by Gene Cluster Alignment. *Syst. Appl. Microbiol.* **21**, 473–477 (1998).
17. Woese, C. R. On the evolution of cells. *Proc. Natl. Acad. Sci.* **99**, 8742–8747 (2002).
18. Miller, S. L. A Production of Amino Acids Under Possible Primitive Earth Conditions. *Science* **117**, 528–529 (1953).
19. Xie, X. *et al.* Primordial soup was edible: abiotically produced Miller-Urey mixture supports bacterial growth. *Sci. Rep.* **5**, 14338 (2015).
20. Jiang, L. *et al.* Abiotic synthesis of amino acids and self-crystallization under prebiotic conditions. *Sci. Rep.* **4**, 6769 (2014).
21. Parker, E. T. *et al.* A Plausible Simultaneous Synthesis of Amino Acids and Simple Peptides on the Primordial Earth. *Angew. Chem. Int. Ed.* **53**, 8132–8136 (2014).
22. Miller, S. L. Production of Some Organic Compounds under Possible Primitive Earth Conditions1. *J. Am. Chem. Soc.* **77**, 2351–2361 (1955).
23. Johnson, A. P. *et al.* The Miller Volcanic Spark Discharge Experiment. *Science* **322**, 404–404 (2008).
24. Bada, J. L. & Lazcano, A. Prebiotic Soup – Revisiting the Miller Experiment. *Science* **300**, 745–746 (2003).
25. Chyba, C. & Sagan, C. Electrical energy sources for organic synthesis on the early earth. *Orig. Life Evol. Biosph.* **21**, 3–17 (1991).

26. Hill, R. D. An efficient lightning energy source on the early earth. *Orig. Life Evol. Biosph.* **22**, 277–285 (1992).
27. Chyba, C. & Sagan, C. Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origins of life. *Nature* **355**, 125–132 (1992).
28. Miller, S. L. The mechanism of synthesis of amino acids by electric discharges. *Biochim. Biophys. Acta* **23**, 480–489 (1957).
29. Wächtershäuser, G. Before enzymes and templates: theory of surface metabolism. *Microbiol. Rev.* **52**, 452–484 (1988).
30. Huber, C. & Wächtershäuser, G. Activated Acetic Acid by Carbon Fixation on (Fe,Ni)S Under Primordial Conditions. *Science* **276**, 245–247 (1997).
31. Wächtershäuser, G. Evolution of the first metabolic cycles. *Proc. Natl. Acad. Sci.* **87**, 200–204 (1990).
32. Wächtershäuser, G. Biomolecules: The origin of their optical activity. *Med. Hypotheses* **36**, 307–311 (1991).
33. Drobner, E., Huber, H., Wächtershäuser, G., Rose, D. & Stetter, K. O. Pyrite formation linked with hydrogen evolution under anaerobic conditions. *Nature* **346**, 742–744 (1990).
34. Huber, C., Eisenreich, W. & Wächtershäuser, G. Synthesis of α-amino and α-hydroxy acids under volcanic conditions: implications for the origin of life. *Tetrahedron Lett.* **51**, 1069–1071 (2010).
35. Huber, C. & Wächtershäuser, G. α-Hydroxy and α-Amino Acids Under Possible Hadean, Volcanic Origin-of-Life Conditions. *Science* **314**, 630–632 (2006).
36. Huber, C., Kraus, F., Hanzlik, M., Eisenreich, W. & Wächtershäuser, G. Elements of Metabolic Evolution. *Chem. – Eur. J.* **18**, 2063–2080 (2012).

37. Keller, M., Blöchl, E., Wächtershäuser, G. & Stetter, K. O. Formation of amide bonds without a condensation agent and implications for origin of life. *Nature* **368**, 836–838 (1994).

38. Huber, C. & Wächtershäuser, G. Peptides by Activation of Amino Acids with CO on (Ni,Fe)S Surfaces: Implications for the Origin of Life. *Science* **281**, 670–672 (1998).

39. Huber, C., Eisenreich, W., Hecht, S. & Wächtershäuser, G. A Possible Primordial Peptide Cycle. *Science* **301**, 938–940 (2003).

40. Lane, N., Allen, J. F. & Martin, W. How did LUCA make a living? Chemiosmosis in the origin of life. *BioEssays* **32**, 271–280 (2010).

41. Jackson, J. B. The "Origin-of-Life Reactor" and Reduction of CO2 by H2 in Inorganic Precipitates. *J. Mol. Evol.* **85**, 1–7 (2017).

42. Jackson, J. B. Ancient Living Organisms Escaping from, or Imprisoned in, the Vents? *Life* **7**, 36 (2017).

43. van der Gulik, P., Massar, S., Gilis, D., Buhrman, H. & Rooman, M. The first peptides: The evolutionary transition between prebiotic amino acids and early proteins. *J. Theor. Biol.* **261**, 531–539 (2009).

44. Orgel, L. E. The Implausibility of Metabolic Cycles on the Prebiotic Earth. *PLoS Biol* **6**, e18 (2008).

45. Anet, F. A. The place of metabolism in the origin of life. *Curr. Opin. Chem. Biol.* **8**, 654–659 (2004).

46. Pascal, R. & Boiteau, L. Energy flows, metabolism and translation. *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 2949–2958 (2011).

47. Schmidt, S., Sunyaev, S., Bork, P. & Dandekar, T. Metabolites: a helping hand for pathway evolution? *Trends Biochem. Sci.* **28**, 336–341 (2003).

48. Ralser, M. The RNA world and the origin of metabolic enzymes. *Biochem. Soc. Trans.* **42**, 985–988 (2014).

49. Hanczyc, M. M., Fujikawa, S. M. & Szostak, J. W. Experimental Models of Primitive Cellular Compartments: Encapsulation, Growth, and Division. *Science* **302**, 618–622 (2003).

50. Adamala, K. & Szostak, J. W. Competition between model protocells driven by an encapsulated catalyst. *Nat. Chem.* **5**, 495–501 (2013).

51. Engelhart, A. E., Adamala, K. P. & Szostak, J. W. A simple physical mechanism enables homeostasis in primitive cells. *Nat. Chem.* **8**, 448–453 (2016).

52. Adamala, K. & Szostak, J. W. Nonenzymatic Template-Directed RNA Synthesis Inside Model Protocells. *Science* **342**, 1098–1100 (2013).

53. Adamala, K. P., Engelhart, A. E. & Szostak, J. W. Collaboration between primitive cell membranes and soluble catalysts. *Nat. Commun.* **7**, 11041 (2016).

54. Cheng, L. K. L. & Unrau, P. J. Closing the Circle: Replicating RNA with RNA. *Cold Spring Harb. Perspect. Biol.* **2**, a002204 (2010).

55. Andras, P. & Andras, C. The origins of life – the 'protein interaction world' hypothesis: protein interactions were the first form of self-reproducing life and nucleic acids evolved later as memory molecules. *Med. Hypotheses* **64**, 678–688 (2005).

56. Orgel, L. E. Some Consequences of the RNA World Hypothesis. *Orig. Life Evol. Biosph.* **33**, 211–218 (2003).

57. Ruiz-Mirazo, K., Briones, C. & de la Escosura, A. Prebiotic Systems Chemistry: New Perspectives for the Origins of Life. *Chem. Rev.* **114**, 285–366 (2014).

58. Ware, E. The Chemistry of the Hydantoins. *Chem. Rev.* **46**, 403–470 (1950).

59. Pascal, R., Taillades, J. & Commeyras, A. Systemes de strecker et apparentes—X: Decomposition et hydratation en milieu aqueux basique des α-aminonitriles secondaires.

Processus d'hydratation autocatalytique et catalyse par l'acetone. *Tetrahedron* **34**, 2275–2281 (1978).

60. Pascal, R., Taillades, J. & Commeyras, A. Systèmes de strecker et apparentes—XII: Catalyse par les aldehydes de l'hydratation intramoleculaire des α-aminonitriles. *Tetrahedron* **36**, 2999–3008 (1980).
61. Taillades, J. *et al.* N-Carbamoyl-α-Amino Acids Rather than Free α-Amino Acids Formation in the Primitive Hydrosphere: A Novel Proposal for the Emergence of Prebiotic Peptides. *Orig. Life Evol. Biosph.* **28**, 61–77 (1998).
62. Chitale, S., Derasp, J. S., Hussain, B., Tanveer, K. & Beauchemin, A. M. Carbohydrates as efficient catalysts for the hydration of α-amino nitriles. *Chem. Commun.* **52**, 13147–13150 (2016).
63. Wagner, A. J., Zubarev, D. Y., Aspuru-Guzik, A. & Blackmond, D. G. Chiral Sugars Drive Enantioenrichment in Prebiotic Amino Acid Synthesis. *ACS Cent. Sci.* **3**, 322–328 (2017).
64. Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D. & Sutherland, J. D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* **7**, 301–307 (2015).
65. Moore, P. B. & Steitz, T. A. The involvement of RNA in ribosome function. *Nature* **418**, 229–235 (2002).
66. Borsenberger, V. *et al.* Exploratory Studies to Investigate a Linked Prebiotic Origin of RNA and Coded Peptides. *Chem. Biodivers.* **1**, 203–246 (2004).
67. van der Gulik, P. T. S. & Speijer, D. How Amino Acids and Peptides Shaped the RNA World. *Life* **5**, 230–246 (2015).
68. Tagami, S., Attwater, J. & Holliger, P. Simple peptides derived from the ribosomal core potentiate RNA polymerase ribozyme function. *Nat. Chem.* **9**, 325–332 (2017).

69. Lee, D. H., Granja, J. R., Martinez, J. A., Severin, K. & Ghadiri, M. R. A self-replicating peptide. *Nature* **382**, 525–528 (1996).

70. Lee, D. H., Severin, K., Yokobayashi, Y. & Ghadiri, M. R. Emergence of symbiosis in peptide self-replication through a hypercyclic network. *Nature* **390**, 591–594 (1997).

71. Rode, B. M. Peptides and the origin of life1. *Peptides* **20**, 773–786 (1999).

72. Paventi, M. & Edward, J. T. Preparation of α-aminothioamides from aldehydes. *Can. J. Chem.* **65**, 282–289 (1987).

73. Leman, L., Orgel, L. & Ghadiri, M. R. Carbonyl Sulfide-Mediated Prebiotic Formation of Peptides. *Science* **306**, 283–286 (2004).

74. Rabinowitz, J., Flores, J., Krebsbach, R. & Rogers, G. Peptide Formation in the Presence of Linear or Cyclic Polyphosphates. *Nature* **224**, 795–796 (1969).

75. Rabinowitz, J. Recherches sur la formation et la transformation des esters LXXXIII [1]. Réactions de condensation et/ou de phosphorylation, en solution aqueuse, de divers composés organiques à fonctions OH, COOH, NH2, ou autres, à l'aide de polyphosphates linéaires ou cycliques. *Helv. Chim. Acta* **52**, 2663–2671 (1969).

76. Ni, F., Gao, X., Zhao, Z.-X., Huang, C. & Zhao, Y.-F. On the Electrophilicity of Cyclic Acylphosphoramidates (CAPAs) Postulated as Intermediates. *Eur. J. Org. Chem.* **2009**, 3026–3035 (2009).

77. Forsythe, J. G. *et al.* Ester-Mediated Amide Bond Formation Driven by Wet–Dry Cycles: A Possible Path to Polypeptides on the Prebiotic Earth. *Angew. Chem. Int. Ed.* **54**, 9871–9875 (2015).

78. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983).

79. Kruger, K. *et al.* Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* **31**, 147–157 (1982).

80. Cech, T. R. The generality of self-splicing RNA: Relationship to nuclear mRNA splicing. *Cell* **44**, 207–210 (1986).

81. Orgel, L. E. Evolution of the genetic apparatus. *J. Mol. Biol.* **38**, 381–393 (1968).

82. Crick, F. H. C. The origin of the genetic code. *J. Mol. Biol.* **38**, 367–379 (1968).

83. Gilbert, W. Origin of life: The RNA world. *Nature* **319**, 618–618 (1986).

84. Robertson, M. P. & Joyce, G. F. The Origins of the RNA World. *Cold Spring Harb. Perspect. Biol.* **4**, a003608 (2012).

85. Cech, T. R. The RNA Worlds in Context. *Cold Spring Harb. Perspect. Biol.* **4**, a006742 (2012).

86. Benner, S. A., Ellington, A. D. & Tauer, A. Modern metabolism as a palimpsest of the RNA world. *Proc. Natl. Acad. Sci.* **86**, 7054–7058 (1989).

87. Stubbe, J., Ge, J. & Yee, C. S. The evolution of ribonucleotide reduction revisited. *Trends Biochem. Sci.* **26**, 93–99 (2001).

88. Caetano-Anollés, G., Kim, H. S. & Mittenthal, J. E. The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. *Proc. Natl. Acad. Sci.* **104**, 9358–9363 (2007).

89. Iii, H. B. W. Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **7**, 101–104 (1976).

90. Neveu, M., Kim, H.-J. & Benner, S. A. The “Strong” RNA World Hypothesis: Fifty Years Old. *Astrobiology* **13**, 391–403 (2013).

91. Sutherland, J. D. Ribonucleotides. *Cold Spring Harb. Perspect. Biol.* **2**, a005439 (2010).

92. Powner, M. W. & Sutherland, J. D. Prebiotic chemistry: a new modus operandi. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 2870–2877 (2011).

93. Copley, S. D., Smith, E. & Morowitz, H. J. The origin of the RNA world: Co-evolution of genes and metabolism. *Bioorganic Chem.* **35**, 430–443 (2007).

94. Eschenmoser, A. Towards a Chemical Etiology of Nucleic Acid Structure. *Orig. Life Evol. Biosph.* **27**, 535–553

95. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science* **289**, 920–930 (2000).

96. Greider, C. W. & Blackburn, E. H. The telomere terminal transferase of tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell* **51**, 887–898 (1987).

97. Hiller, D. A. & Strobel, S. A. The chemical versatility of RNA. *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 2929–2935 (2011).

98. Krupkin, M. *et al.* A vestige of a prebiotic bonding machine is functioning within the contemporary ribosome. *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 2972–2978 (2011).

99. Lilley, D. M. J. Mechanisms of RNA catalysis. *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 2910–2917 (2011).

100. Moore, P. B. & Steitz, T. A. The Structural Basis of Large Ribosomal Subunit Function. *Annu. Rev. Biochem.* **72**, 813–850 (2003).

101. Polacek, N., Gaynor, M., Yassin, A. & Mankin, A. S. Ribosomal peptidyl transferase can withstand mutations at the putative catalytic nucleotide. *Nature* **411**, 498–501 (2001).

102. Gibson, T. J. & Lamond, A. I. Metabolic complexity in the RNA world and implications for the origin of protein synthesis. *J. Mol. Evol.* **30**, 7–15 (1990).

103. Ekland, E. H. & Bartel, D. P. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **382**, 373–376 (1996).

104. Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E. & Bartel, D. P. RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension. *Science* **292**, 1319–1325 (2001).

105. McGinness, K. E. & Joyce, G. F. RNA-Catalyzed RNA Ligation on an External RNA Template. *Chem. Biol.* **9**, 297–307 (2002).

106. Unrau, P. J. & Bartel, D. P. RNA-catalysed nucleotide synthesis. *Nature* **395**, 260–263 (1998).

107. Kumar, R. K. & Yarus, M. RNA-Catalyzed Amino Acid Activation†. *Biochemistry* **40**, 6998–7004 (2001).

108. Suga, H., Hayashi, G. & Terasaka, N. The RNA origin of transfer RNA aminoacylation and beyond. *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 2959–2964 (2011).

109. Tarasow, T. M., Tarasow, S. L. & Eaton, B. E. RNA-catalysed carbon–carbon bond formation. *Nature* **389**, 54–57 (1997).

110. Yarus, M. The meaning of a minuscule ribozyme. *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 2902–2909 (2011).

111. Verlander, M. S., Lohrmann, R. & Orgel, L. E. Catalysts for the self-polymerization of adenosine cyclic 2',3'-phosphate. *J. Mol. Evol.* **2**, 303–316 (1973).

112. Verlander, M. S. & Orgel, L. E. Analysis of high molecular weight material from the polymerization of adenosine cyclic 2', 3'-phosphate. *J. Mol. Evol.* **3**, 115–120 (1974).

113. Rohatgi, R., Bartel, D. P. & Szostak, J. W. Kinetic and Mechanistic Analysis of Nonenzymatic, Template-Directed Oligoribonucleotide Ligation. *J. Am. Chem. Soc.* **118**, 3332–3339 (1996).

114. Orgel, L. E. Molecular replication. *Nature* **358**, 203–209 (1992).

115. Zhang, W., Tam, C. P., Wang, J. & Szostak, J. W. Unusual Base-Pairing Interactions in Monomer–Template Complexes. *ACS Cent. Sci.* **2**, 916–926 (2016).

116. Rajamani, S. *et al.* Effect of Stalling after Mismatches on the Error Catastrophe in Nonenzymatic Nucleic Acid Replication. *J. Am. Chem. Soc.* **132**, 5880–5885 (2010).

117. Kreysing, M., Keil, L., Lanzmich, S. & Braun, D. Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length. *Nat. Chem.* **7**, 203–208 (2015).

118. Saladino, R., Botta, G., Pino, S., Costanzo, G. & Mauro, E. D. Genetics first or metabolism first? The formamide clue. *Chem. Soc. Rev.* **41**, 5526–5565 (2012).

119. Islam, S. & Powner, M. W. Prebiotic Systems Chemistry: Complexity Overcoming Clutter. *Chem* **2**, 470–501 (2017).

120. Crowe, M. A. & Sutherland, J. D. Reaction of Cytidine Nucleotides with Cyanoacetylene: Support for the Intermediacy of Nucleoside-2′,3′-cyclic Phosphates in the Prebiotic Synthesis of RNA. *ChemBioChem* **7**, 951–956 (2006).

121. Renz, M., Lohrmann, R. & Orgel, L. E. Catalysts for the polymerization of adenosine cyclic 2′,3′-phosphate on a poly (U) template. *Biochim. Biophys. Acta BBA - Nucleic Acids Protein Synth.* **240**, 463–471 (1971).

122. James, K. D. & Ellington, A. D. The Fidelity of Template-Directed Oligonucleotide Ligation and the Inevitability of Polymerase Function. *Orig. Life Evol. Biosph.* **29**, 375–390 (1999).

123. Vázquez-Salazar, A. *et al.* Can an Imidazole Be Formed from an Alanyl-Seryl-Glycine Tripeptide under Possible Prebiotic Conditions? *Orig. Life Evol. Biospheres* **47**, 345–354 (2017).

124. Jauker, M., Griesser, H. & Richert, C. Copying of RNA Sequences without Pre-Activation. *Angew. Chem. Int. Ed.* **54**, 14559–14563 (2015).

125. Blain, J. C. & Szostak, J. W. Progress Toward Synthetic Cells. *Annu. Rev. Biochem.* **83**, 615–640 (2014).

126. Wu, T. & Orgel, L. E. Nonenzymic template-directed synthesis on hairpin oligonucleotides. 2. Templates containing cytidine and guanosine residues. *J. Am. Chem. Soc.* **114**, 5496–5501 (1992).

127. Pal, A. *et al.* Effect of terminal 3′-hydroxymethyl modification of an RNA primer on nonenzymatic primer extension. *Chem. Commun.* **52**, 11905–11907 (2016).

128. Xu, J. *et al.* A prebiotically plausible synthesis of pyrimidine β-ribonucleosides and their phosphate derivatives involving photoanomerization. *Nat. Chem.* **9**, 301–309 (2017).

129. Izgu, E. C., Oh, S. S. & Szostak, J. W. Synthesis of activated 3'-amino-3'-deoxy-2-thio-thymidine, a superior substrate for the nonenzymatic copying of nucleic acid templates. *Chem. Commun.* **52**, 3684–3686 (2016).

130. Heuberger, B. D., Pal, A., Del Frate, F., Topkar, V. V. & Szostak, J. W. Replacing Uridine with 2-Thiouridine Enhances the Rate and Fidelity of Nonenzymatic RNA Primer Extension. *J. Am. Chem. Soc.* **137**, 2769–2775 (2015).

131. Prywes, N., Michaels, Y. S., Pal, A., Oh, S. S. & Szostak, J. W. Thiolated uridine substrates and templates improve the rate and fidelity of ribozyme-catalyzed RNA copying. *Chem. Commun.* **52**, 6529–6532 (2016).

132. Walton, T. & Szostak, J. W. A Highly Reactive Imidazolium-Bridged Dinucleotide Intermediate in Nonenzymatic RNA Primer Extension. *J. Am. Chem. Soc.* **138**, 11996–12002 (2016).

133. Tam, C. P. *et al.* Downstream Oligonucleotides Strongly Enhance the Affinity of GMP to RNA Primer–Template Complexes. *J. Am. Chem. Soc.* **139**, 571–574 (2017).

134. Prywes, N., Blain, J. C., Frate, F. D. & Szostak, J. W. Nonenzymatic copying of RNA templates containing all four letters is catalyzed by activated oligonucleotides. *eLife* **5**, e17756 (2016).

135. Zhang, W. *et al.* Insight into the mechanism of nonenzymatic RNA primer extension from the structure of an RNA-GpppG complex. *Proc. Natl. Acad. Sci.* **114**, 7659–7664 (2017).

136. Li, L. *et al.* Enhanced Nonenzymatic RNA Copying with 2-Aminoimidazole Activated Nucleotides. *J. Am. Chem. Soc.* **139**, 1810–1813 (2017).

137. Fahrenbach, A. C. *et al.* Common and Potentially Prebiotic Origin for Precursors of Nucleotide Synthesis and Activation. *J. Am. Chem. Soc.* **139**, 8780–8783 (2017).

138. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242 (2009).

139. Bowler, F. R. *et al.* Prebiotically plausible oligoribonucleotide ligation facilitated by chemoselective acetylation. *Nat. Chem.* **5**, 383–389 (2013).

140. Mariani, A. & Sutherland, J. D. Non-Enzymatic RNA Backbone Proofreading through Energy-Dissipative Recycling. *Angew. Chem. Int. Ed.* **56**, 6563–6566 (2017).

141. Sheng, J. *et al.* Structural insights into the effects of 2'-5' linkages on the RNA duplex. *Proc. Natl. Acad. Sci.* **111**, 3050–3055 (2014).

142. Engelhart, A. E., Powner, M. W. & Szostak, J. W. Functional RNAs exhibit tolerance for non-heritable 2'–5' versus 3'–5' backbone heterogeneity. *Nat. Chem.* **5**, 390–394 (2013).

143. Jones, B. L. On the evolution of prebiotic enzyme coupled macromolecular systems. *Bull. Math. Biol.* **42**, 539–549 (1980).

144. Hulshof, J. & Ponnamperuma, C. Prebiotic condensation reactions in an aqueous medium: A review of condensing agents. *Orig. Life* **7**, 197–224 (1976).

145. Butlerow, A. Bildung einer zuckerartigen Substanz durch Synthese. *Justus Liebigs Ann. Chem.* **120**, 295–298

146. Breslow, R. On the mechanism of the formose reaction. *Tetrahedron Lett.* **1**, 22–26 (1959).

147. Ricardo, A., Carrigan, M. A., Olcott, A. N. & Benner, S. A. Borate Minerals Stabilize Ribose. *Science* **303**, 196–196 (2004).

148. Kim, H.-J. *et al.* Synthesis of Carbohydrates in Mineral-Guided Prebiotic Cycles. *J. Am. Chem. Soc.* **133**, 9457–9468 (2011).

149. Appayee, C. & Breslow, R. Deuterium Studies Reveal a New Mechanism for the Formose Reaction Involving Hydride Shifts. *J. Am. Chem. Soc.* **136**, 3720–3723 (2014).

150. Shapiro, R. Prebiotic ribose synthesis: A critical analysis. *Orig. Life Evol. Biosph.* **18**, 71–85 (1988).

151. Müller, D. *et al.* Chemie von a-Aminonitrilen. Aldomerisierung von Glycolaldehyd-phosphat zu racemischen Hexose-2,4,6-triphosphaten und (in Gegenwart von Formaldehyd) racemischen Pentose-2,4-diphosphaten: rac-Allose-2,4,6-triphosphat und rac-Ribose-2,4-diphosphat sind die Reaktionshauptprodukte. *Helv. Chim. Acta* **73**, 1410–1468 (1990).

152. Krishnamurthy, R., Arrhenius, G. & Eschenmoser, A. Formation of Glycolaldehyde Phosphate from Glycolaldehyde in Aqueous Solution. *Orig. Life Evol. Biosph.* **29**, 333–354

153. Krishnamurthy, R., Guntha, S. & Eschenmoser, A. Regioselective α-Phosphorylation of Aldoses in Aqueous Solution. *Angew. Chem. Int. Ed.* **39**, 2281–2285 (2000).

154. Coggins, A. J. & Powner, M. W. Prebiotic synthesis of phosphoenol pyruvate by α-phosphorylation-controlled triose glycolysis. *Nat. Chem.* **9**, 310–317 (2017).

155. Krishnamurthy, R., Pitsch, S. & Arrhenius, G. Mineral Induced Formation of Pentose-2,4-Bisphosphates*. *Orig. Life Evol. Biosph.* **29**, 139–152 (1999).

156. Usami, K. & Okamoto, A. Hydroxyapatite: catalyst for a one-pot pentose formation. *Org. Biomol. Chem.* **15**, 8888–8893 (2017).

157. Tewari, Y. B. & Goldberg, R. N. An investigation of the equilibria between aqueous ribose, ribulose, and arabinose. *Biophys. Chem.* **22**, 197–204 (1985).

158. Larralde, R., Robertson, M. P. & Miller, S. L. Rates of Decomposition of Ribose and Other Sugars: Implications for Chemical Evolution. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 8158–8160 (1995).

159. Furukawa, Y., Horiuchi, M. & Kakegawa, T. Selective Stabilization of Ribose by Borate. *Orig. Life Evol. Biospheres* **43**, 353–361 (2013).

160. Scorei, R. & Cimpoiaşu, V. M. Boron Enhances the Thermostability of Carbohydrates. *Orig. Life Evol. Biospheres* **36**, 1–11 (2006).

161. Scorei, R. Is Boron a Prebiotic Element? A Mini-review of the Essentiality of Boron for the Appearance of Life on Earth. *Orig. Life Evol. Biospheres* **42**, 3–17 (2012).

162. Oró, J. Synthesis of adenine from ammonium cyanide. *Biochem. Biophys. Res. Commun.* **2**, 407–412 (1960).

163. Oró, J. Mechanism of Synthesis of Adenine from Hydrogen Cyanide under Possible Primitive Earth Conditions. *Nature* **191**, 1193–1194 (1961).

164. Shapiro, R. The prebiotic role of adenine: A critical analysis. *Orig. Life Evol. Biosph.* **25**, 83–98 (1995).

165. Ferris, J. P. & Orgel, L. E. Studies in Prebiotic Synthesis. I. Aminomalononitrile and 4-Amino-5-cyanoimidazole1,2. *J. Am. Chem. Soc.* **88**, 3829–3831 (1966).

166. Sanchez, R. A., Ferbis, J. P. & Orgel, L. E. Studies in Prebiodc Synthesis: II. Synthesis of purine precursors and amino acids from aqueous hydrogen cyanide. *J. Mol. Biol.* **30**, 223–253 (1967).

167. Sanchez, R. A., Ferris, J. P. & Orgel, L. E. Studies in prebiotic synthesis: IV. Conversion of 4-aminoimidazole-5-carbonitrile derivatives to purines. *J. Mol. Biol.* **38**, 121–128 (1968).

168. Sanchez, R., Ferris, J. & Orgel, L. E. Conditions for Purine Synthesis: Did Prebiotic Synthesis Occur at Low Temperatures? *Science* **153**, 72–73 (1966).

169. Orgel, L. E. Prebiotic Adenine Revisited: Eutectics and Photochemistry. *Orig. Life Evol. Biosph.* **34**, 361–369 (2004).

170. Ferris, J. P., Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis: III. Synthesis of pyrimidines from cyanoacetylene and cyanate. *J. Mol. Biol.* **33**, 693–704 (1968).

171. Robertson, M. P. & Miller, S. L. An efficient prebiotic synthesis of cytosine and uracil. *Nature* **375**, 772–774 (1995).

172. Fuller, W. D., Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis: VI. Synthesis of purine nucleosides. *J. Mol. Biol.* **67**, 25–33 (1972).

173. Prieur, B. E. Étude de l'activité prébiotique potentielle de l'acide borique. *Comptes Rendus Académie Sci. - Ser. IIC - Chem.* **4**, 667–670 (2001).

174. Zubay, G. & Mui, T. Prebiotic Synthesis of Nucleotides. *Orig. Life Evol. Biosph.* **31**, 87–102 (2001).

175. Kim, H.-J. & Benner, S. A. Prebiotic stereoselective synthesis of purine and noncanonical pyrimidine nucleotide from nucleobases and phosphorylated carbohydrates. *Proc. Natl. Acad. Sci.* **114**, 11315–11320 (2017).

176. Westheimer, F. H. Why Nature Chose Phosphates. *Science* **235**, 1173–1178 (1987).

177. Davis, B. D. On the importance of being ionized. *Arch. Biochem. Biophys.* **78**, 497–509 (1958).

178. Kamerlin, S. C. L., Sharma, P. K., Prasad, R. B. & Warshel, A. Why nature really chose phosphate. *Q. Rev. Biophys.* **46**, 1–132 (2013).

179. Griffith, E. J., Ponnamperuma, C. & Gabel, N. W. Phosphorus, a key to life on the primitive earth. *Orig. Life* **8**, 71–85 (1977).

180. Maciá*, E., Hernández, M. V. & Oró, J. Primary Sources of Phosphorus and Phosphates in Chemical Evolution. *Orig. Life Evol. Biosph.* **27**, 459–480 (1997).

181. GULICK, A. PHOSPHORUS AS A FACTOR IN THE ORIGIN OF LIFE. *Am. Sci.* **43**, 479–489 (1955).

182. Keefe, A. D. & Miller, S. L. Are polyphosphates or phosphate esters prebiotic reagents? *J. Mol. Evol.* **41**, 693–702 (1995).

183. Pasek, M. A. Rethinking early Earth phosphorus geochemistry. *Proc. Natl. Acad. Sci.* **105**, 853–858 (2008).

184. Pasek, M. A. & Lauretta, D. S. Aqueous Corrosion of Phosphide Minerals from Iron Meteorites: A Highly Reactive Source of Prebiotic Phosphorus on the Surface of the Early Earth. *Astrobiology* **5**, 515–535 (2005).

185. Pasek, M. A., Harnmeijer, J. P., Buick, R., Gull, M. & Atlas, Z. Evidence for reactive reduced phosphorus species in the early Archean ocean. *Proc. Natl. Acad. Sci.* **110**, 10089–10094 (2013).

186. Pasek, M. A., Dworkin, J. P. & Lauretta, D. S. A radical pathway for organic phosphorylation during schreibersite corrosion with implications for the origin of life. *Geochim. Cosmochim. Acta* **71**, 1721–1736 (2007).

187. Yamagata, Y., Watanabe, H., Saitoh, M. & Namba, T. Volcanic production of polyphosphates and its relevance to prebiotic evolution. *Nature* **352**, 516–519 (1991).

188. Ferris, J. P. Cyanovinyl Phosphate: A Prebiological Phosphorylating Agent? *Science* **161**, 53–54 (1968).

189. Lohrmann, R. & Orgel, L. E. Prebiotic Synthesis: Phosphorylation in Aqueous Solution. *Science* **161**, 64–66 (1968).

190. Halmann, M., Sanchez, R. A. & Orgel, L. E. Phosphorylation of D-ribose in aqueous solution. *J. Org. Chem.* **34**, 3702–3703 (1969).

191. Nam, I., Lee, J. K., Nam, H. G. & Zare, R. N. Abiotic production of sugar phosphates and uridine ribonucleoside in aqueous microdroplets. *Proc. Natl. Acad. Sci.* **114**, 12396–12400 (2017).

192. Ponnamperuma, C. & Mack, R. Nucleotide Synthesis under Possible Primitive Earth Conditions. *Science* **148**, 1221–1223 (1965).

193. Beck, A., Lohrmann, R. & Orgel, L. E. Phosphorylation with Inorganic Phosphates at Moderate Temperatures. *Science* **157**, 952–952 (1967).

194. Handschuh, G. J., Lohrmann, R. & Orgel, L. E. The effect of Mg2+ and Ca2+ on urea-catalyzed phosphorylation reactions. *J. Mol. Evol.* **2**, 251–262 (1973).

195. Lohrmann, R. & Orgel, L. E. Urea-Inorganic Phosphate Mixtures as Prebiotic Phosphorylating Agents. *Science* **171**, 490–494 (1971).

196. Burcar, B. *et al.* Darwin's Warm Little Pond: A One-Pot Reaction for Prebiotic Phosphorylation and the Mobilization of Phosphate from Minerals in a Urea-Based Solvent. *Angew. Chem. Int. Ed.* **55**, 13249–13253 (2016).

197. Handschuh, G. J. & Orgel, L. E. Struvite and Prebiotic Phosphorylation. *Science* **179**, 483–484 (1973).

198. Gull, M. & Pasek, M. A. Is Struvite a Prebiotic Mineral? *Life* **3**, 321–330 (2013).

199. Kim, H.-J. *et al.* Evaporite Borate-Containing Mineral Ensembles Make Phosphate Available and Regiospecifically Phosphorylate Ribonucleosides: Borate as a Multifaceted Problem Solver in Prebiotic Chemistry. *Angew. Chem. Int. Ed.* **55**, 15816–15820 (2016).

200. Pasek, M. A., Kee, T. P., Bryant, D. E., Pavlov, A. A. & Lunine, J. I. Production of Potentially Prebiotic Condensed Phosphates by Phosphorus Redox Chemistry. *Angew. Chem. Int. Ed.* **47**, 7918–7920

201. Osterberg, R. & Orgel, L. E. Polyphosphate and trimetaphosphate formation under potentially prebiotic conditions. *J. Mol. Evol.* **1**, 241–248 (1972).

202. Schwartz, A. W. Specific phosphorylation of the 2'- and 3'- positions in ribonucleosides. *J. Chem. Soc. Chem. Commun.* **0**, 1393a-1393a (1969).

203. Saffhill, R. Selective phosphorylation of the cis-2', 3'-diol of unprotected ribonucleosides with trimetaphosphate in aqueous solution. *J. Org. Chem.* **35**, 2881–2883 (1970).

204. Yamagata, Y., Inoue, H. & Inomata, K. Specific effect of magnesium ion on 2′, 3′-cyclic amp synthesis from adenosine and trimeta phosphate in aqueous solution. *Orig. Life Evol. Biosph.* **25**, 47–52 (1995).

205. Tsuhako, M., Fujimoto, M., Ohashi, S., Nariai, H. & Motooka, I. Phosphorylation of Nucleosides with Sodium cyclo-Triphosphate. *Bull. Chem. Soc. Jpn.* **57**, 3274–3280 (1984).

206. Quimby, O. T. & Flautt, T. J. Ammonolyse des Trimetaphosphats. *Z. Für Anorg. Allg. Chem.* **296**, 220–228

207. Feldmann, W. & Thilo, E. Zur Chemie der kondensierten Phosphate und Arsenate. XXXVIII. Amidotriphosphat. *Z. Für Anorg. Allg. Chem.* **328**, 113–126

208. Gibard, C., Bhowmik, S., Karki, M., Kim, E.-K. & Krishnamurthy, R. Phosphorylation, oligomerization and self-assembly in water under potential prebiotic conditions. *Nat. Chem.* **10**, 212–217 (2017).

209. Sacerdote, M. G. & Szostak, J. W. Semipermeable lipid bilayers exhibit diastereoselectivity favoring ribose. *Proc. Natl. Acad. Sci.* **102**, 6004–6008 (2005).

210. Eschenmoser, A. Chemical Etiology of Nucleic Acid Structure. *Science* **284**, 2118–2124 (1999).

211. The quest for the chemical roots of life. *Chem. Commun.* **0**, 1247–1252 (2004).

212. Eschenmoser, A. & Dobler, M. Warum Pentose- und nicht Hexose-Nucleinsäuren?? Teil I. Einleitung und Problemstellung, Konformationsanalyse für Oligonucleotid-Ketten aus 2′,3′-Dideoxyglucopyranosyl-Bausteinen (‘Homo-DNS’) sowie Betrachtungen zur Konformation von A- und B-DNS. *Helv. Chim. Acta* **75**, 218–259 (1992).

213. Böhringer, M. *et al.* Warum Pentose- und nicht Hexose-Nucleinsäuren??. Teil II. Oligonucleotide aus 2′,3′-Dideoxy-β-D-glucopyranosyl-Bausteinen (‘Homo-DNS’): Herstellung. *Helv. Chim. Acta* **75**, 1416–1477 (1992).

214. Hunziker, J. *et al.* Warum pentose-und nicht hexose-nucleinsäuren? Teil III. Oligo(2',3'-dideoxy-β-D-glucopyranosyl) nucleotide ('homo-DNS'): Paarungesigenschaften. *Helv. Chim. Acta* **76**, 259–352 (1993).

215. Pitsch, S., Wendeborn, S., Jaun, B. & Eschenmoser, A. Why Pentose- and Not Hexose-Nucleic Acids??. Part VII. Pyranosyl-RNA ('p-RNA'). Preliminary communication. *Helv. Chim. Acta* **76**, 2161–2183 (1993).

216. Egli, M. *et al.* Crystal Structure of Homo-DNA and Nature's Choice of Pentose over Hexose in the Genetic System. *J. Am. Chem. Soc.* **128**, 10847–10856 (2006).

217. Egli, M., Lubini, P. & Pallan, P. S. The long and winding road to the structure of homo-DNA. *Chem. Soc. Rev.* **36**, 31–45 (2006).

218. Otting, G. *et al.* Warum Pentose- und nicht Hexose-Nucleinsäuren??. Teil VI. 'Homo-DNS': 1H-, 13C-, 31P- und 15N-NMR-spektroskopische Untersuchung von ddGlc(A-A-A-A-A-T-T-T-T-T) in wässriger Lösung. *Helv. Chim. Acta* **76**, 2701–2756 (1993).

219. Pitsch, S. *et al.* Pyranosyl-RNA ('p-RNA'): Base-pairing selectivity and potential to replicate. Preliminary communication. *Helv. Chim. Acta* **78**, 1621–1635 (1995).

220. Beier, M., Reck, F., Wagner, T., Krishnamurthy, R. & Eschenmoser, A. Chemical Etiology of Nucleic Acid Structure: Comparing Pentopyranosyl-(2'→4') Oligonucleotides with RNA. *Science* **283**, 699–703 (1999).

221. Reck, F. *et al.* l-α-Lyxopyranosyl (4'→3') Oligonucleotides: A Base-Pairing System Containing a Shortened Backbone1. *Org. Lett.* **1**, 1531–1534 (1999).

222. Ichida, J. K., Horhota, A., Zou, K., McLaughlin, L. W. & Szostak, J. W. High fidelity TNA synthesis by Therminator polymerase. *Nucleic Acids Res.* **33**, 5219–5225 (2005).

223. Schöning, K.-U. *et al.* Chemical Etiology of Nucleic Acid Structure: The α-Threofuranosyl-(3'→2') Oligonucleotide System. *Science* **290**, 1347–1351 (2000).

224. Pallan, P. S. *et al.* Why Does TNA Cross-Pair More Strongly with RNA Than with DNA? An Answer From X-ray Analysis. *Angew. Chem. Int. Ed.* **42**, 5893–5895 (2003).

225. Wu, X., Guntha, S., Ferencic, M., Krishnamurthy, R. & Eschenmoser, A. Base-Pairing Systems Related to TNA: α-Threofuranosyl Oligonucleotides Containing Phosphoramidate Linkages1. *Org. Lett.* **4**, 1279–1282 (2002).

226. Herdewijn, P. TNA as a Potential Alternative to Natural Nucleic Acids. *Angew. Chem. Int. Ed.* **40**, 2249–2251 (2001).

227. Zhang, L., Peritz, A. & Meggers, E. A Simple Glycol Nucleic Acid. *J. Am. Chem. Soc.* **127**, 4174–4175 (2005).

228. Zhang, L., Peritz, A. E., Carroll, P. J. & Meggers, E. Synthesis of Glycol Nucleic Acids. *Synthesis* **2006**, 645–653 (2006).

229. Meggers, E. & Zhang, L. Synthesis and Properties of the Simplified Nucleic Acid Glycol Nucleic Acid. *Acc. Chem. Res.* **43**, 1092–1102 (2010).

230. Cooper, G. *et al.* Carbonaceous meteorites as a source of sugar-related organic compounds for the early Earth. *Nature* **414**, 879–883 (2001).

231. Karri, P., Punna, V., Kim, K. & Krishnamurthy, R. Base-Pairing Properties of a Structural Isomer of Glycerol Nucleic Acid. *Angew. Chem. Int. Ed.* **52**, 5840–5844 (2013).

232. Krishnamurthy, R. RNA as an Emergent Entity: An Understanding Gained Through Studying its Nonfunctional Alternatives. *Synlett* **25**, 1511–1517 (2014).

233. Benner, S. A., Yang, Z. & Chen, F. Synthetic biology, tinkering biology, and artificial biology. What are we learning? *Comptes Rendus Chim.* **14**, 372–387 (2011).

234. Joyce, G. F., Schwartz, A. W., Miller, S. L. & Orgel, L. E. The case for an ancestral genetic system involving simple analogues of the nucleotides. *Proc. Natl. Acad. Sci.* **84**, 4398–4402 (1987).

235. Krishnamurthy, R. On the Emergence of RNA. *Isr. J. Chem.* **55**, 837–850 (2015).

236. Krishnamurthy, R. Giving Rise to Life: Transition from Prebiotic Chemistry to Protobiology. *Acc. Chem. Res.* **50**, 455–459 (2017).

237. Wittung, P., Nielsen, P. E., Buchardt, O., Egholm, M. & Norde´n, B. DNA-like double helix formed by peptide nucleic acid. *Nature* **368**, 561–563 (1994).

238. Egholm, M., Buchardt, O., Nielsen, P. E. & Berg, R. H. Peptide nucleic acids (PNA). Oligonucleotide analogs with an achiral peptide backbone. *J. Am. Chem. Soc.* **114**, 1895–1897 (1992).

239. Egholm, M. *et al.* PNA hybridizes to complementary oligonucleotides obeying the Watson–Crick hydrogen-bonding rules. *Nature* **365**, 566–568 (1993).

240. Schmidt, J. G., Christensen, L., Nielsen, P. E. & Orgel, L. E. Information transfer from DNA to peptide nucleic acids by template-directed syntheses. *Nucleic Acids Res.* **25**, 4792–4796 (1997).

241. Hutter, D., Blaettler, M. O. & Benner, S. A. From Phosphate to Bis(methylene) Sulfone: Non-Ionic Backbone Linkers in DNA. *Helv. Chim. Acta* **85**, 2777–2806 (2002).

242. Benner, S. A. Understanding Nucleic Acids Using Synthetic Chemistry. *Acc. Chem. Res.* **37**, 784–797 (2004).

243. Hud, N. V., Cafferty, B. J., Krishnamurthy, R. & Williams, L. D. The Origin of RNA and "My Grandfather's Axe". *Chem. Biol.* **20**, 466–474 (2013).

244. Krishnamurthy, R. Role of pKa of Nucleobases in the Origins of Chemical Evolution. *Acc. Chem. Res.* **45**, 2035–2044 (2012).

245. Luisi, P. L. Emergence in Chemistry: Chemistry as the Embodiment of Emergence. *Found. Chem.* **4**, 183–200 (2002).

246. Nelson, K. E., Levy, M. & Miller, S. L. Peptide nucleic acids rather than RNA may have been the first genetic molecule. *Proc. Natl. Acad. Sci.* **97**, 3868–3871 (2000).

247. Wächtershäuser, G. An All-Purine Precursor of Nucleic Acids. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 1134–1135 (1988).

248. Nelsestuen, G. L. Origin of life: Consideration of alternatives to proteins and nucleic acids. *J. Mol. Evol.* **15**, 59–72 (1980).

249. Cairns-Smith, A. G. The origin of life and the nature of the primitive gene. *J. Theor. Biol.* **10**, 53–88 (1966).

250. Powner, M. W., Zheng, S.-L. & Szostak, J. W. Multicomponent Assembly of Proposed DNA Precursors in Water. *J. Am. Chem. Soc.* **134**, 13889–13895 (2012).

251. Anastasi, C. *et al.* RNA: Prebiotic Product, or Biotic Invention? *Chem. Biodivers.* **4**, 721–739 (2007).

252. Springsteen, G. & Joyce, G. F. Selective Derivatization and Sequestration of Ribose from a Prebiotic Mix. *J. Am. Chem. Soc.* **126**, 9578–9583 (2004).

253. Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis: V. Synthesis and photoanomerization of pyrimidine nucleosides. *J. Mol. Biol.* **47**, 531–543 (1970).

254. Powner, M. W. *et al.* On the Prebiotic Synthesis of Ribonucleotides: Photoanomerisation of Cytosine Nucleosides and Nucleotides Revisited. *ChemBioChem* **8**, 1170–1179 (2007).

255. Powner, M. W. & Sutherland, J. D. Potentially Prebiotic Synthesis of Pyrimidine β-D-Ribonucleotides by Photoanomerization/Hydrolysis of α-D-Cytidine-2′-Phosphate. *ChemBioChem* **9**, 2386–2387 (2008).

256. Tapiero, C. M. & Nagyvary, J. Prebiotic Formation of Cytidine Nucleotides. *Nature* **231**, 42–43 (1971).

257. Ingar, A.-A., Luke, R. W. A., Hayter, B. R. & Sutherland, J. D. Synthesis of Cytidine Ribonucleotides by Stepwise Assembly of the Heterocycle on a Sugar Phosphate. *ChemBioChem* **4**, 504–507 (2003).

258. Cockerill, A. F. *et al.* An Improved Synthesis of 2-Amino-1,3-oxazoles under Basic Catalysis. *Synthesis* **1976**, 591–593 (1976).

259. Anastasi, C., Crowe, M. A., Powner, M. W. & Sutherland, J. D. Direct Assembly of Nucleoside Precursors from Two- and Three-Carbon Units. *Angew. Chem. Int. Ed.* **45**, 6176–6179 (2006).

260. Choudhary, A., Kamer, K. J., Powner, M. W., Sutherland, J. D. & Raines, R. T. A Stereoelectronic Effect in Prebiotic Nucleotide Synthesis. *ACS Chem. Biol.* **5**, 655–657 (2010).

261. Anastasi, C., Crowe, M. A. & Sutherland, J. D. Two-Step Potentially Prebiotic Synthesis of α-d-Cytidine-5′-phosphate from d-Glyceraldehyde-3-phosphate. *J. Am. Chem. Soc.* **129**, 24–25 (2006).

262. Powner, M. W. & Sutherland, J. D. Phosphate-Mediated Interconversion of Ribo- and Arabino-Configured Prebiotic Nucleotide Intermediates. *Angew. Chem. Int. Ed.* **49**, 4641–4643 (2010).

263. Ludlow, R. F. & Otto, S. Systems chemistry. *Chem. Soc. Rev.* **37**, 101–108 (2007).

264. Nitschke, J. R. Systems chemistry: Molecular networks come of age. *Nature* (2009). doi:10.1038/462736a

265. Peretó, J. Out of fuzzy chemistry: from prebiotic chemistry to metabolic networks. *Chem. Soc. Rev.* **41**, 5394–5403 (2012).

266. Islam, S., Bučar, D.-K. & Powner, M. W. Prebiotic selection and assembly of proteinogenic amino acids and natural nucleotides from complex mixtures. *Nat. Chem.* **9**, 584–589 (2017).

267. Ritson, D. & Sutherland, J. D. Prebiotic synthesis of simple sugars by photoredox systems chemistry. *Nat. Chem.* **4**, 895–899 (2012).

268. Ritson, D. J. & Sutherland, J. D. Synthesis of Aldehydic Ribonucleotide and Amino Acid Precursors by Photoredox Chemistry. *Angew. Chem. Int. Ed.* **52**, 5845–5847 (2013).

269. Ritson, D. J. & Sutherland, J. D. Conversion of Biosynthetic Precursors of RNA to Those of DNA by Photoredox Chemistry. *J. Mol. Evol.* **78**, 245–250 (2014).

270. Taillades, J. & Commeyras, A. Systemes de strecker et apparentes—III: Etude en solution aqueuse de la stabilite et des conditions de synthese des α-aminonitriles tertiaires. Importance des protons portes par le groupement amine. *Tetrahedron* **30**, 3407–3414 (1974).

271. Lundberg, H., Tinnis, F., Selander, N. & Adolfsson, H. Catalytic amide formation from non-activated carboxylic acids and amines. *Chem. Soc. Rev.* **43**, 2714–2742 (2014).

272. Jakubke, H.-D., Kuhl, P. & Könnecke, A. Basic Principles of Protease-Catalyzed Peptide Bond Formation. *Angew. Chem. Int. Ed. Engl.* **24**, 85–93

273. Liu, R. & Orgel, L. E. Polymerization of β-amino Acids in Aqueous Solution. *Orig. Life Evol. Biosph.* **28**, 47–60 (1998).

274. Novelli, G. D. Amino Acid Activation for Protein Synthesis. *Annu. Rev. Biochem.* **36**, 449–484 (1967).

275. Rabinowitz, J. Note on the rôle of cyanides and polyphosphates in the formation of peptides in aqueous solutions of amino acids, at room temperature, as a possible prebiotic process. *Helv. Chim. Acta* **54**, 1483–1485 (1971).

276. Rabinowitz, J. Peptide and Amide Bond Formation in Aqueous Solutions of Cyclic or Linear Polyphosphates as a possible prebiotic process. *Helv. Chim. Acta* **53**, 1350–1355 (1970).

277. Chung, N. M., Lohrmann, R., Orgel, L. E. & Rabinowitz, J. The mechanism of the trimetaphosphate-induced peptide synthesis. *Tetrahedron* **27**, 1205–1210 (1971).

278. Dhiman, R. S., Opinska, L. G. & Kluger, R. Biomimetic peptide bond formation in water with aminoacyl phosphate esters. *Org. Biomol. Chem.* **9**, 5645–5647 (2011).

279. Ni, F., Sun, S., Huang, C. & Zhao, Y. N-phosphorylation of amino acids by trimetaphosphate in aqueous solution—learning from prebiotic synthesis. *Green Chem.* **11**, 569–573 (2009).

280. Paecht, M. & Katchalsky, A. Aspartic acid formation from glycine phosphate. *Biochim. Biophys. Acta BBA - Gen. Subj.* **90**, 260–281 (1964).

281. Koshland, D. E. Kinetics of Peptide Bond Formation1. *J. Am. Chem. Soc.* **73**, 4103–4108 (1951).

282. Tsuhako, M. *et al.* The Reaction of cyclo-Triphosphate with L-α- or β-Alanine. *Bull. Chem. Soc. Jpn.* **58**, 3092–3098 (1985).

283. Inoue, H., Baba, Y., Furukawa, T., Maeda, Y. & Tsuhako, M. Formation of Dipeptide in the Reaction of Amino Acids with cyclo-Triphosphate. *Chem. Pharm. Bull. (Tokyo)* **41**, 1895–1899 (1993).

284. Gorenstein, D. G. Dependence of phosphorus-31 chemical shifts on oxygen-phosphorus-oxygen bond angles in phosphate esters. (2002). doi:10.1021/ja00837a039

285. Gorenstein, D. G. & Kar, D. 31P chemical shifts in phosphate diester monoanions. Bond angle and torsional angle effects. *Biochem. Biophys. Res. Commun.* **65**, 1073–1080 (1975).

286. Anastasi, C. *et al.* The Search for a Potentially Prebiotic Synthesis of Nucleotides via Arabinose-3-phosphate and Its Cyanamide Derivative. *Chem. – Eur. J.* **14**, 2375–2388 (2008).

287. Eschenmoser, A. & Loewenthal, E. Chemistry of potentially prebiological natural products. *Chem. Soc. Rev.* **21**, 1–16 (1992).

288. Jauker, M., Griesser, H. & Richert, C. Spontaneous Formation of RNA Strands, Peptidyl RNA, and Cofactors. *Angew. Chem. Int. Ed.* **54**, 14564–14569 (2015).

289. Halmann, M. Cyanamide-induced condensation reactions of glycine. *Arch. Biochem. Biophys.* **128**, 808–810 (1968).

290. Danger G. *et al.* 5(4H)-Oxazolones as Intermediates in the Carbodiimide- and Cyanamide-Promoted Peptide Activations in Aqueous Solution. *Angew. Chem. Int. Ed.* **52**, 611–614 (2012).

291. Danger, G., Boiteau, L., Cottet, H. & Pascal, R. The Peptide Formation Mediated by Cyanate Revisited. N-Carboxyanhydrides as Accessible Intermediates in the Decomposition of N-Carbamoylamino Acids. *J. Am. Chem. Soc.* **128**, 7412–7413 (2006).

292. Leman, L. J., Orgel, L. E. & Ghadiri, M. R. Amino Acid Dependent Formation of Phosphate Anhydrides in Water Mediated by Carbonyl Sulfide. *J. Am. Chem. Soc.* **128**, 20–21 (2006).

293. Biron, J.-P. & Pascal, R. Amino Acid N-Carboxyanhydrides: Activated Peptide Monomers Behaving as Phosphate-Activating Agents in Aqueous Solution. *J. Am. Chem. Soc.* **126**, 9198–9199 (2004).

294. Paecht, M. & Katchalsky, A. Phosphorylation of Adenylic Acid by the Phosphate Anhydride of Leucine and Chromatographic Analysis of the Resulting Products. *J. Am. Chem. Soc.* **76**, 6197–6198 (1954).

295. Lassila, J. K., Zalatan, J. G. & Herschlag, D. Biological Phosphoryl-Transfer Reactions: Understanding Mechanism and Catalysis. *Annu. Rev. Biochem.* **80**, 669–702 (2011).

296. Biron, J.-P., Parkes, A. L., Pascal, R. & Sutherland, J. D. Expeditious, Potentially Primordial, Aminoacylation of Nucleotides. *Angew. Chem. Int. Ed.* **44**, 6731–6734 (2005).

297. Wickramasinghe, N. S. M. D., Staves, M. P. & Lacey, J. C. Stereoselective, nonenzymic, intramolecular transfer of amino acids. *Biochemistry* **30**, 2768–2772 (1991).

298. Izgu, E. C. *et al.* N-Carboxyanhydride-Mediated Fatty Acylation of Amino Acids and Peptides for Functionalization of Protocell Membranes. *J. Am. Chem. Soc.* **138**, 16669–16676 (2016).

299. Leman, L. J. & Ghadiri, M. R. Potentially Prebiotic Synthesis of α-Amino Thioacids in Water. *Synlett* **28**, 68–72 (2017).

300. Leman, L. J., Huang, Z.-Z. & Ghadiri, M. R. Peptide Bond Formation in Water Mediated by Carbon Disulfide. *Astrobiology* **15**, 709–716 (2015).

301. Maurel, M. C. & Orgel, L. E. Oligomerization of alpha-thioglutamic acid. *Orig. Life Evol. Biosphere J. Int. Soc. Study Orig. Life* **30**, 423–430 (2000).

302. Liu, R. & Orgel, L. E. Oxidative acylation using thioacids. *Nature* **389**, 52–54 (1997).

303. Weber, A. L. & Orgel, L. E. The formation of peptides from glycine thioesters. *J. Mol. Evol.* **13**, 193–202 (1979).

304. Weber, A. L. Prebiotic amino acid thioester synthesis: thiol-dependent amino acid synthesis from formose substrates (formaldehyde and glycolaldehyde) and ammonia. *Orig. Life Evol. Biosphere J. Int. Soc. Study Orig. Life* **28**, 259–270 (1998).

305. Chadha, M. S., Replogle, L., Flores, J. & Ponnamperuma, C. Possible role of aminoacetonitrile in chemical evolution. *Bioorganic Chem.* **1**, 269–274 (1971).

306. Cook, A. H., Heilbron, I. & Mahadevan, A. P. 225. Studies in the azole series. Part XI. The interaction of α-amino-nitriles, hydrogen sulphide, and ketones. *J. Chem. Soc. Resumed* **0**, 1061–1064 (1949).

307. Taillades, J. & Commeyras, A. Systemes de strecker et apparentes—II: Mécanisme de formation en solution aqueuse des α-alcoylaminoisobutyronitrile à partir d'acétone, d'acide cyanhydrique et d'ammoniaque, methyl ou diméthylamine. *Tetrahedron* **30**, 2493–2501 (1974).

308. Stewart, R. & Van Dyke, J. D. The Hydration of Ketones in Mixtures of Water and Polar Aprotic Solvents. *Can. J. Chem.* **50**, 1992–1999 (1972).

309. Cocker, W. & Lapworth, A. CLXXXVII.—Experiments on the synthetic preparation and isolation of some of the simpler amino-acids. *J. Chem. Soc. Resumed* **0**, 1391–1403 (1931).

310. Mowry, D. T. The Preparation of Nitriles. *Chem. Rev.* **42**, 189–283 (1948).

311. Wolman, Y., Haverland, W. J. & Miller, S. L. Nonprotein Amino Acids from Spark Discharges and Their Comparison with the Murchison Meteorite Amino Acids. *Proc. Natl. Acad. Sci.* **69**, 809–811 (1972).

312. Schlesinger, G. & Miller, S. L. Equilibrium and kinetics of glyconitrile formation in aqueous solution. *J. Am. Chem. Soc.* **95**, 3729–3735 (1973).

313. Rossi Jean-Christophe, Marull Marc, Boiteau Laurent & Taillades Jacques. Regioselective Hydration and Deprotection of Chiral, Dissymmetric Iminodinitriles in the Scope of an Asymmetric Strecker Strategy. *Eur. J. Org. Chem.* **2007**, 662–668 (2006).

314. Bahde, R. J. & Rychnovsky, S. D. Cyclization via Carbolithiation of α-Amino Alkyllithium Reagents. *Org. Lett.* **10**, 4017–4020 (2008).

315. Béjaud, M., Mion, L., Taillades, J. & Commeyras, A. Etude comparative de la reactivite des α-aminonitriles secondaires et tertiaires en solution aqueuse entre pH 10 et 14. Hydrolyse des α-aminonitriles secondaires et son importance dans la formation prebiotique des acides amines naturels. *Tetrahedron* **31**, 403–410 (1975).

316. Filipe, H. A. L. *et al.* Synthesis and Characterization of a Lipidic Alpha Amino Acid: Solubility and Interaction with Serum Albumin and Lipid Bilayers. (2013). doi:10.1021/jp307874v

317. Strulson, C. A., Molden, R. C., Keating, C. D. & Bevilacqua, P. C. RNA catalysis through compartmentalization. *Nat. Chem.* **4**, 941–946 (2012).

318. Kamat, N. P., Tobé, S., Hill, I. T. & Szostak, J. W. Electrostatic Localization of RNA to Protocell Membranes by Cationic Hydrophobic Peptides. *Angew. Chem. Int. Ed.* **54**, 11735–11739 (2015).

319. Trump, J. E. V. & Miller, S. L. Prebiotic Synthesis of Methionine. *Science* **178**, 859–860 (1972).

320. Fernández-García, C., Grefenstette, N. M. & Powner, M. W. Prebiotic synthesis of aminooxazoline-5'-phosphates in water by oxidative phosphorylation. *Chem. Commun.* **53**, 4919–4921 (2017).

321. Cook, A. H. & Cox, S. F. 494. The preparation of α-N-alkylamino-nitriles, -amides, and -acids. *J. Chem. Soc. Resumed* **0**, 2334–2337 (1949).

322. Hasan Misbah Ul. 13C NMR spectra of some amides and imides. Effect of inductive and mesomeric interactions, cyclization and hydrogen bonding on 13C NMR chemical shifts. *Org. Magn. Reson.* **14**, 447–450 (2005).

323. Johnson, T. B. & Burnham, G. Thioamides: The Formation of Thiopolypeptide Derivatives by the Action of Hydrogen Sulphide on Aminoacetonitrile. *J. Biol. Chem.* **9**, 449–462 (1911).

324. Benner, S. A., Kim, H.-J. & Carrigan, M. A. Asphalt, Water, and the Prebiotic Synthesis of Ribose, Ribonucleosides, and RNA. *Acc. Chem. Res.* **45**, 2025–2034 (2012).

325. Benner, S. A., Kim, H.-J. & Yang, Z. Setting the Stage: The History, Chemistry, and Geobiology behind RNA. *Cold Spring Harb. Perspect. Biol.* **4**, (2012).

326. Eschenmoser, A. Etiology of Potentially Primordial Biomolecular Structures: From Vitamin B12 to the Nucleic Acids and an Inquiry into the Chemistry of Life's Origin: A Retrospective. *Angew. Chem. Int. Ed.* **50**, 12412–12472 (2011).

327. Sheng, Y., Bean, H. D., Mamajanov, I., Hud, N. V. & Leszczynski, J. Comprehensive Investigation of the Energetics of Pyrimidine Nucleoside Formation in a Model Prebiotic Reaction. *J. Am. Chem. Soc.* **131**, 16088–16095 (2009).

328. Schwartz, A. W. Phosphorus in prebiotic chemistry. *Philos. Trans. R. Soc. B Biol. Sci.* **361**, 1743–1749 (2006).

329. Fernández-García, C., Coggins, A. J. & Powner, M. W. A Chemist's Perspective on the Role of Phosphorus at the Origins of Life. *Life* **7**, (2017).

330. Martin, A. R., Vasseur, J.-J. & Smietana, M. Boron and nucleic acid chemistries: merging the best of both worlds. *Chem. Soc. Rev.* **42**, 5684–5713 (2013).

331. Schubert, D. Boron Oxides, Boric Acid, and Borates. in *Kirk-Othmer Encyclopedia of Chemical Technology* (John Wiley & Sons, Inc., 2000). doi:10.1002/0471238961.0215181519130920.a01.pub3

332. Grew, E. S., Bada, J. L. & Hazen, R. M. Borate Minerals and Origin of the RNA World. *Orig. Life Evol. Biospheres* **41**, 307–316 (2011).

333. Kakegawa, T., Noda, M. & Nannri, H. Geochemical Cycles of Bio-essential Elements on the Early Earth and Their Relationships to Origin of Life. *Resour. Geol.* **52**, 83–89 (2002).

334. Martin, A. R., Barvik, I., Luvino, D., Smietana, M. & Vasseur, J.-J. Dynamic and Programmable DNA-Templated Boronic Ester Formation. *Angew. Chem. Int. Ed.* **50**, 4193–4196 (2011).

335. Luvino, D., Smietana, M. & Vasseur, J.-J. Selective fluorescence-based detection of dihydrouridine with boronic acids. *Tetrahedron Lett.* **47**, 9253–9256 (2006).

336. Kim, D. H., Marbois, B. N., Faull, K. F. & Eckhert, C. D. Esterification of borate with NAD+ and NADH as studied by electrospray ionization mass spectrometry and 11B NMR spectroscopy. *J. Mass Spectrom.* **38**, 632–640 (2003).

337. Kim, D. H., Faull, K. F., Norris, A. J. & Eckhert, C. D. Borate–nucleotide complex formation depends on charge and phosphorylation state. *J. Mass Spectrom.* **39**, 743–751 (2004).

338. Martin, A. R. *et al.* Expanding the bononucleotide family: synthesis of borono-analogues of dCMP, dGMP and dAMP. *Org. Biomol. Chem.* **7**, 4369–4377 (2009).

339. Luvino, D., Baraguey, C., Smietana, M. & Vasseur, J.-J. Bononucleotides: synthesis, and formation of a new reversible boronate internucleosidic linkage. *Chem. Commun.* 2352–2354 (2008). doi:10.1039/B802098A

340. Saladino, R., Barontini, M., Cossetti, C., Mauro, E. D. & Crestini, C. The Effects of Borate Minerals on the Synthesis of Nucleic Acid Bases, Amino Acids and Biogenic Carboxylic Acids from Formamide. *Orig. Life Evol. Biospheres* **41**, 317–330 (2011).

341. Chapelle, S. & Verchere, J.-F. A 11b and 13c nmr determination of the structures of borate complexes of pentoses and related sugars. *Tetrahedron* **44**, 4469–4482 (1988).

342. Stairs, S. *et al.* Divergent prebiotic synthesis of pyrimidine and 8-oxo-purine ribonucleotides. *Nat. Commun.* **8**, 15270 (2017).

343. McKillop, A. & Sanderson, W. R. Sodium perborate and sodium percarbonate: Cheap, safe and versatile oxidising agents for organic synthesis. *Tetrahedron* **51**, 6145–6166 (1995).

344. McKillop, A. & Kemp, D. Further functional group oxidations using sodium perborate. *Tetrahedron* **45**, 3299–3306 (1989).

345. Chattopadhyaya, J. B. & Reese, C. B. Interconversion of 8,2'-O-cycloadenosine and 2',3'-anhydro-8-oxyadenosine. *J. Chem. Soc. Chem. Commun.* 860–862 (1976). doi:10.1039/C39760000860

346. Heather D. Bean, David G. Lynn & Nicholas V. Hud. Self-Assembly and the Origin of the First RNA-Like Polymers. in *Chemical Evolution II: From the Origins of Life to Modern Society* **1025**, 109–132 (American Chemical Society, 2009).

347. Weber, B. H. On the Emergence of Living Systems. *Biosemiotics* **2**, 343–359 (2009).

348. Fuller, W. D., Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis. VII. *J. Mol. Evol.* **1**, 249–257 (1972).

349. Bean, H. D. *et al.* Formation of a β-Pyrimidine Nucleoside by a Free Pyrimidine Base and Ribose in a Plausible Prebiotic Reaction. *J. Am. Chem. Soc.* **129**, 9556–9557 (2007).

350. Saladino, R. *et al.* Proton irradiation: a key to the challenge of N-glycosidic bond formation in a prebiotic context. *Sci. Rep.* **7**, 14709 (2017).

351. Kolb, V. M., Dworkin, J. P. & Miller, S. L. Alternative bases in the RNA world: The prebiotic synthesis of urazole and its ribosides. *J. Mol. Evol.* **38**, 549–557 (1994).

352. Kim, H.-J. & Benner, S. A. Prebiotic Glycosylation of Uracil with Electron-Donating Substituents. *Astrobiology* **15**, 301–306 (2015).

353. Chen, M. C. *et al.* Spontaneous Prebiotic Formation of a β-Ribofuranoside That Self-Assembles with a Complementary Heterocycle. *J. Am. Chem. Soc.* **136**, 5640–5646 (2014).

354. Cafferty, B. J., Fialho, D. M., Khanam, J., Krishnamurthy, R. & Hud, N. V. Spontaneous formation and base pairing of plausible prebiotic nucleotides in water. *Nat. Commun.* **7**, 11328 (2016).

355. Fernández-García, C., Grefenstette, N. M. & Powner, M. W. Selective aqueous acetylation controls the photoanomerization of α-cytidine-5′-phosphate. *Chem. Commun.* **54**, 4850–4853 (2018).

356. Johnson, T. B. & Hilbert, G. E. The Synthesis of Pyrimidine-Nucleosides. *Science* **69**, 579–580 (1929).

357. Hilbert, G. E. & Johnson, T. B. RESEARCHES ON PYRIMIDINES. CXVII. A METHOD FOR THE SYNTHESIS OF NUCLEOSIDES. *J. Am. Chem. Soc.* **52**, 4489–4494 (1930).

358. Fernández-García, C. & Powner, M. W. Selective Acylation of Nucleosides, Nucleotides, and Glycerol-3-phosphocholine in Water. *Synlett* **28**, 78–83 (2017).

359. Ness, R. K., Diehl, H. W. & Fletcher, H. G. New Benzoyl Derivatives of D-Ribofuranose and aldehydo-D-Ribose. The Preparation of Crystalline 2,3,5-Tri-O-benzoyl-β-D-ribose from D-Ribose1. *J. Am. Chem. Soc.* **76**, 763–767 (1954).

360. Schwögler, A., Gramlich, V. & Carell, T. Synthesis and Properties of Flavin Ribofuranosides and Flavin Ribopyranosides. *Helv. Chim. Acta* **83**, 2452–2463 (2000).

361. Hennen, W. J., Sweers, H. M., Wang, Y. F. & Wong, C. H. Enzymes in carbohydrate synthesis. Lipase-catalyzed selective acylation and deacylation of furanose and pyranose derivatives. *J. Org. Chem.* **53**, 4939–4945 (1988).

362. Nudelman, A., Herzig, J., Gottlieb, H. E., Keinan, E. & Sterling, J. Selective deacetylation of anomeric sugar acetates with tin alkoxides. *Carbohydr. Res.* **162**, 145–152 (1987).

363. Gravier-Pelletier, C., Ginisty, M. & Le Merrer, Y. A versatile scaffold for a library of liposidomycins analogues: a crucial and potent glycosylation step. *Tetrahedron Asymmetry* **15**, 189–193 (2004).

364. Becker, S. *et al.* A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science* **352**, 833–836 (2016).

365. Powner, M. W., Sutherland, J. D. & Szostak, J. W. Chemoselective Multicomponent One-Pot Assembly of Purine Precursors in Water. *J. Am. Chem. Soc.* **132**, 16677–16688 (2010).

366. Ikehara, M. & Ogiso, Y. Studies of nucleosides and nucleotides—54 : Purine cyclonucleosides—19. Further investigations on the cleavage of the 8,2′-O-anhydro

linkage. A new synthesis of 9-β-D-arabinofuranosyladenine. *Tetrahedron* **28**, 3695–3704 (1972).

367. Ikehara, M. & Kaneko, M. A novel method for the synthesis of cyclonucleosides. Synthesis of 8,5'-0-anhydro-8-oxyadenosine. *J. Am. Chem. Soc.* **90**, 497–498 (1968).

368. Nagpal, K. L. & Dhar, M. M. The use of a purine cyclonucleoside for the synthesis of a dinucleoside phosphate. *Tetrahedron Lett.* **9**, 47–49 (1968).

369. Sako, M., Saito, T., Kameyama, K., Hirota, K. & Maki, Y. Reductive cleavage of the O–C(8) bond in 5'-O,8-cycloadenosines. Intramolecular protection of the 8-position and the 5'-hydroxy group in adenosines. *J. Chem. Soc. Chem. Commun.* 1298–1299 (1987). doi:10.1039/C39870001298

370. Coulter, C. L. Structural chemistry of cyclic nucleotides. II. Crystal and molecular structure of sodium .beta.-cytidine 2',3'-cyclic phosphate. *J. Am. Chem. Soc.* **95**, 570–575 (1973).

371. Abraham, R. J., Hall, L. D., Hough, L. & McLauchlan, K. A. 717. A proton resonance study of the conformations of carbohydrates in solution. Part I. Derivatives of 1,2-O-isopropylidene-α-D-xylo-hexofuranose. *J. Chem. Soc. Resumed* **0**, 3699–3705 (1962).

372. Powner, M., Sutherland, J. & Szostak, J. The Origins of Nucleotides. *Synlett* **2011**, 1956–1964 (2011).

373. Reist, E. J., Benitez, A. & Goodman, L. The Synthesis of Some 5'-Thiopentofuranosylpyrimidines1. *J. Org. Chem.* **29**, 554–558 (1964).

374. Santi, D. V. & Brewer, C. F. Model studies of thymidylate synthetase. Neighboring-group facilitation of electrophilic substitution reactions of uracil furanosides. *J. Am. Chem. Soc.* **90**, 6236–6238 (1968).

375. Otter, B. A., Falco, E. A. & Fox, J. J. Nucleosides. LVIII. Transformations of pyrimidine nucleosides in alkaline media. 3. Conversion of 5-halouridines into imidazoline and barbituric acid nucleosides. *J. Org. Chem.* **34**, 1390–1396 (1969).

376. Altona, C. & Sundaralingam, M. Conformational analysis of the sugar ring in nucleosides and nucleotides. New description using the concept of pseudorotation. (2002). doi:10.1021/ja00778a043

377. Neidle, S. 2 - The Building-Blocks of DNA and RNA. in *Principles of Nucleic Acid Structure* 20–37 (Academic Press, 2008). doi:10.1016/B978-012369507-9.50003-0

378. Maki, Y., Sako, M., Saito, T. & Hirota, K. Intramolecular Cyclization of Purine Nucleosides by N-Halogenosuccinimides/Acetic Acid. A Mechanistic Aspect on the C(8)-Halogenation of Purine Nucleosides. *HETEROCYCLES* **27**, 347 (1988).

379. Qu, G.-R., Ren, B., Niu, H.-Y., Mao, Z.-J. & Guo, H.-M. A Novel One-Step Method for the Synthesis of C-5-Substituted O6,5′-Cyclopyrimidine Nucleoside Analogues in Ammonia Water. *J. Org. Chem.* **73**, 2450–2453 (2008).

380. Ikehara, M., Uesigi, S. & Kaneko, M. Bromination of Adenine Nucleotide and Nucleoside. *Chem. Commun.* 17–18 (1967).

381. Ikehara, M., Kaneko, M. & Okano, R. Studies of nucleosides and nucleotides—45: Purine cyclonucleosides.-12. Synthesis of adenine cyclonucleosides having 8,5′-0-anhydro linkage. *Tetrahedron* **26**, 5675–5682 (1970).

382. Swarbrick, J. M. & Potter, B. V. L. Total Synthesis of a Cyclic Adenosine 5′-Diphosphate Ribose Receptor Agonist. *J. Org. Chem.* **77**, 4191–4197 (2012).

383. Ikehara, M., Kaneko, M. & Sagai, M. Synthesis of 8, 5′-Anhydro-8-mercaptoadenosine. *Chem. Pharm. Bull. (Tokyo)* **16**, 1151–1153 (1968).

384. Choi, S. *et al.* Oxidation of Guanosine Derivatives by a Platinum(IV) Complex: Internal Electron Transfer through Cyclization. *J. Am. Chem. Soc.* **127**, 1773–1781 (2005).

385. Kohda, K., Baba, K. & Kawazoe, Y. Synthesis of 7-Aminoguanosine and Its Conversion to 8-Substituted Derivatives. *Chem. Pharm. Bull. (Tokyo)* **34**, 2298–2301 (1986).

386. Guengerich, F. P., Mundkowski, R. G., Voehler, M. & Kadlubar, F. F. Formation and Reactions of N7-Aminoguanosine and Derivatives. *Chem. Res. Toxicol.* **12**, 906–916 (1999).

387. Gani, D. & Johnson, A. W. The base-sugar conformation of certain derivatives of adenosine. *J. Chem. Soc. Perkin 1* 1197–1204 (1982). doi:10.1039/P19820001197

388. IKEHARA, M., KANEKO, M., NAKAHARA, Y., YAMADA, S. & UESUGI, S. Studies of Nucleosides and Nucleotides. 43. Purine Cyclonucleosides. (10). Optical Rotatory Dispersion and Circular Dichroism of Adenine 8-Cyclonucleosides. *Chem. Pharm. Bull. (Tokyo)* **19**, 1381–1388 (1971).

389. NAKA, T. & HONJO, M. Synthesis of 8-Carbamoyl- and 8-Carboxyadenosine 3', 5'-Cyclic Phosphates. *Chem. Pharm. Bull. (Tokyo)* **24**, 2052–2056 (1976).

390. Ikehara, M., Tada, H. & Kaneko, M. Studies of nucleosides and nucleotides—35: Purine Cyclonucleosides—5. Synthesis of purine cyclonucleoside having 8,2'-O-anhydro linkage and its cleavage reactions. *Tetrahedron* **24**, 3489–3498 (1968).

391. Ikehara, M. Purine 8-cyclonucleosides. *Acc. Chem. Res.* **2**, 47–53 (1969).

392. Yu, M., Wang, Z., Hu, J., Li, S. & Du, H. Copper-Catalyzed Intramolecular Alkoxylation of Purine Nucleosides: One-Step Synthesis of 5'-O,8-Cyclopurine Nucleosides. *J. Org. Chem.* **80**, 9446–9453 (2015).

393. Capon, R. J. & Trotter, N. S. N3,5'-Cycloxanthosine, the First Natural Occurrence of a Cyclonucleoside. *J. Nat. Prod.* **68**, 1689–1691 (2005).

394. Karpeisky, A., Zavgorodny, S., Hotokka, M., Oivanen, M. & Lönnberg, H. Kinetics for the acid-catalysed hydrolysis of O-, S- and N-bridged 5',8-cyclonucleosides related to adenosine. *J. Chem. Soc. Perkin Trans. 2* 741–744 (1994). doi:10.1039/P29940000741

395. Framski, G., Gdaniec, Z., Gdaniec, M. & Boryski, J. A reinvestigated mechanism of ribosylation of adenine under silylating conditions. *Tetrahedron* **62**, 10123–10129 (2006).

396. Herskovits, T. T. & Harrington, J. P. Solution studies of the nucleic acid bases and related model compounds. Solubility in aqueous alcohol and glycol solutions. *Biochemistry* **11**, 4800–4811 (1972).

397. Choi, S. *et al.* Mechanism of Two-Electron Oxidation of Deoxyguanosine 5'-Monophosphate by a Platinum(IV) Complex. *J. Am. Chem. Soc.* **126**, 591–598 (2004).

398. Choi, S. *et al.* Oxidation of a guanine derivative coordinated to a Pt(IV) complex initiated by intermolecular nucleophilic attacks. *Dalton Trans.* **40**, 2888–2897 (2011).

399. Ritson, D. J., Battilocchio, C., Ley, S. V. & Sutherland, J. D. Mimicking the surface and prebiotic chemistry of early Earth using flow chemistry. *Nat. Commun.* **9**, 1821 (2018).

400. McClendon, J. H. Elemental abundance as a factor in the origins of mineral nutrient requirements. *J. Mol. Evol.* **8**, 175–195 (1976).

401. Wächtershäuser, G. In Praise of Error. *J. Mol. Evol.* **82**, 75–80 (2016).

402. Li, Y. *et al.* A New Radical Scavenging Anthracene Glycoside, Asperflavin Ribofuranoside, and Polyketides from a Marine Isolate of the Fungus Microsporum. *Chem. Pharm. Bull. (Tokyo)* **54**, 882–883 (2006).

403. Pahnke, K. & Meier, C. Synthesis of a Bioreversibly Masked Lipophilic Adenosine Diphosphate Ribose Derivative. *ChemBioChem* **18**, 1616–1626 (2017).

404. Jawalekar, A. M. *et al.* Conjugation of Nucleosides and Oligonucleotides by [3+2] Cycloaddition. *J. Org. Chem.* **73**, 287–290 (2008).

405. Uesugi, S., Tanaka, S. & Ikehara, M. Carbon-13 Nuclear-Magnetic-Resonance Spectra of Adenine Cyclonucleosides and Their Phosphates. *Eur. J. Biochem.* **90**, 205–212 (1978).

406. Ikehara, M. & Uesugi, S. Studies of Nucleosides and Nucleotides. 38. Synthesis of 8-Bromoadenosine Nucleotides. *Chem. Pharm. Bull. (Tokyo)* **17**, 348–354 (1969).

407. Ikehara, M. & Muneyama, K. Studies of Nucleosides and Nucleotides. 40. Synthesis of 8, 5'-O-Cyclonucleoside derived from 8-Oxyguanosine and Its Cleavage by Nucleophilic Reagents. *Chem. Pharm. Bull. (Tokyo)* **18**, 1196–1200 (1970).

408. Nilov, D. I. *et al.* Oxidation of Adenosine and Inosine: The Chemistry of 8-Oxo-7,8-dihydropurines, Purine Iminoquinones, and Purine Quinones as Observed by Ultrafast Spectroscopy. *J. Am. Chem. Soc.* **135**, 3423–3438 (2013).

409. Strohmeier, J., Hertel, I., Diederichsen, U., Rudolph, M. G. & Klostermeier, D. Changing nucleotide specificity of the DEAD-box helicase Hera abrogates communication between the Q-motif and the P-loop. *Biol. Chem.* **392**, 357–369 (2011).

410. Sekine, M. *et al.* Synthesis of Chemically Stabilized Phosmidosine Analogues and the Structure−Activity Relationship of Phosmidosine. *J. Org. Chem.* **69**, 314–326 (2004).

411. Phillips, D. R., Uramoto, M., Isono, K. & McCloskey, J. A. Structure of the antifungal nucleotide antibiotic phosmidosine. *J. Org. Chem.* **58**, 854–859 (1993).

412. Bernadou, J., Gelas, P. & Meunier, B. Hydroxylating activity of a water-soluble manganese porphyrin associated with potassium hydrogen persulfate: Formation of 8-hydroxyadenosine-5'-mono-phosphate from AMP. *Tetrahedron Lett.* **29**, 6615–6617 (1988).

413. Saladino, R., Crestini, C., Bernini, R., Mincione, E. & Ciafrino, R. A new and efficient synthesis of 8-hydroxypurine derivatives by dimethyldioxirane oxidation. *Tetrahedron Lett.* **36**, 2665–2668 (1995).

414. Uesugi, S. & Ikehara, M. Studies of Nucleosides and Nucleotides. 81. : Carbon-13 Magnetic Resonance Spectra of 8-Substituted Purine Nucleotides. Effects of Various Phosphate Groups on the Chemical Shifts and Conformation of Nucleotides. *Chem. Pharm. Bull. (Tokyo)* **26**, 3040–3049 (1978).

415. Wyrick, S. D. & Chaney, S. G. Synthesis of [195mPt]-tetraplatin. *J. Label. Compd. Radiopharm.* **28**, 753–756 (1990).